

BEST AVAILABLE COPY

J. Biol. Chem. 264, 14005-14010, 1989

A:Title: Intersubunit fluorescence energy transfer in human factor VIII.
A:Reference number: A43986; MUID:89340500
A:Accession: A43986
A:Molecule type: protein
A:Residues: 17, 517-523;1853-1860,'X',1862-1864,'X',1866 <FAY>
R:Leylet, A.; van Schijndel, H.B.; Niehrs, C.; Hutter, W.B.; Verbeet, M.P.; Mertens, K.;
J. Biol. Chem. 266, 740-746, 1991
A:Title: Sulfation of Tyr(1860) of human blood coagulation factor VIII is essential for
A:Reference number: A56109
A:Contents: annotation; sulfation
R:Galschier, J.; Wood, W.I.; Gosalika, T.M.; Wlton, K.L.; Chen, E.Y.; Eaton, D.H.; Vehar,
Nature 312, 326-330, 1984
A:Title: Characterization of the human factor VIII gene.
A:Reference number: A56196; MUID:85061547
A:Contents: annotation; introns
R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M.
Protein Sci. 4, 740-746, 1995
A:Title: Locations of disulfide bonds and free cysteines in the heavy and light chains of
A:Reference number: A56216; MUID:95338127
A:Contents: annotation; disulfide bonds
R:Kjake, M.; Hedner, U.; Nalbo, G.; Persson, E.; Thomsen, J.; Ezban, M.
Eur. J. Biochem. 234, 773-779, 1995
A:Title: Amino acid residues 721-729 are required for full factor VIII activity.
A:Reference number: S63527; MUID:96163459
A:Accession: S63527
A:Molecule type: protein
A:Residues: 733-752;753-759 <KJA>
R:Lind, P.; Larsson, K.; Spira, J.; Sydow-Baeckman, M.; Almstedt, A.; Gray, E.; Sandberg
Eur. J. Biochem. 232, 19-27, 1995
A:Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction
A:Reference number: S66445; MUID:96048024
A:Accession: S66445
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1668-1685 <LIN>
C:Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure to
C:Genetics:
A:Gene: GDB:F8C
A:Cross-references: GDB:119124; OMIM:306700
A:Map position: Xq28-Xq28
A:Intons: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 63
C:Function:
A:Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa pro-
A:Pathway: blood coagulation
A:Family: coagulation factor VIII; discooidin I amino-terminal homology; ferroxidase
A:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla-
1/Domain: signal sequence #status predicted <SIC>
F:20-2351/Product: coagulation factor VIII #status experimental <MAT>
F:20-356/Domain: A1 <DA1>
F:23-448/Domain: ferroxidase repeat homology <FOI>
F:392-759/Domain: A2 <DA2>
F:402-730/Domain: ferroxidase repeat homology <FOI>
F:760-1667/Domain: B <DBO>
F:1668-2351/Product: coagulation factor VIIIA light chain #status experimental <ACL>
F:1709-2038/Domain: A3 <DA3>
F:1716-2038/Domain: ferroxidase repeat homology <FO3>
F:2039-2191/Domain: C1 <DC1>
F:2039-2188/Domain: discooidin I amino-terminal homology <DNI>
F:2192-2351/Domain: C2 <DC2>
F:2192-2351/Domain: discooidin I amino-terminal homology <DNI>
F:2192-2351/Domain: discooidin I amino-terminal homology <DNI>
F:60,258,601,776,803,847,919,962,982,1020,1024,1074,1085,1204,1274,1278,1301,1319,1403,1
F:172-198,267-348,547-573,649-730,1851-1877,1918-1922,2040-2188/Disulfide bonds: #status
F:355-356/Cleavage site: Arg-Met (coagulation factor Xa, protein C) #status predicted
F:365,737,738,742,1683,1699/Binding site: sulfite (Tyr) (covalent) #status experimental
F:391-392/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F:414,426/Binding site: sulfite (Tyr) (covalent) #status predicted
F:759-760/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F:1667-1668/Cleavage site: Arg-Glu (unidentified proteinase) #status experimental
F:1708-1709/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F:1740-1741/Cleavage site: Arg-Ala (coagulation factor Xa) #status experimental
F:2193-2345/Disulfide bonds: #status predicted

Query Match 94.1%; Score 7340; DB 1; Length 2351;
Best Local Similarity 62.0%; Pred. No. 0;
Matches 1457; Conservative 0; Mismatches 0; Indels 894; Gaps 1;
QY 1 MQELSTCFLLRFLRFSATRRYLYCAVELSWDYSQSDGELPVDARFPFRPKSPFN 60
Db 1 MQELSTCFLLRFLRFSATRRYLYCAVELSWDYSQSDGELPVDARFPFRPKSPFN 60
QY 61 TSVYKRTLPFEPDHLFNIAKPPPMGLPGLTOAEVDTVITLKNASHPVSLHAV 120
Db 61 TSVYKRTLPFEPDHLFNIAKPPPMGLPGLTOAEVDTVITLKNASHPVSLHAV 120
QY 121 GVSYSKASEGAEYDQTSOREKEDKVPFGSHYVQVYKENGPMASDPLCLTYSLSH 180
Db 121 GVSYSKASEGAEYDQTSOREKEDKVPFGSHYVQVYKENGPMASDPLCLTYSLSH 180
QY 181 VDLVKDLSGLIGALLVCREGSLAKERTQTLHKFILLFAVDEKSWHSETKNSLMODRD 240
Db 181 VDLVKDLSGLIGALLVCREGSLAKERTQTLHKFILLFAVDEKSWHSETKNSLMODRD 240
QY 241 AASARAPKMTYNGYVNRSLPGLGHRKSVYHVGMTPEVHSIFLEGHTFLVRNH 300
Db 241 AASARAPKMTYNGYVNRSLPGLGHRKSVYHVGMTPEVHSIFLEGHTFLVRNH 300
QY 301 ROASLEISPIITFLTAQTLMDLGOFLFCHISSHQHDGMEAYVYVDSCEPEPOLRKNNE 360
Db 301 ROASLEISPIITFLTAQTLMDLGOFLFCHISSHQHDGMEAYVYVDSCEPEPOLRKNNE 360
QY 361 EAEDYDDDLTSEMDYVAFDDNSPSFIQISVAKKHPKTVVHTIAEEDMDVAPLVIA 420
Db 361 EAEDYDDDLTSEMDYVAFDDNSPSFIQISVAKKHPKTVVHTIAEEDMDVAPLVIA 420
QY 421 PDRSRYKSOYLNNGPORIGRKYKRYKRYMAVDETFKTRALQHSGLIGPLLYEGVDTL 480
Db 421 PDRSRYKSOYLNNGPORIGRKYKRYKRYMAVDETFKTRALQHSGLIGPLLYEGVDTL 480
QY 481 LIIFKQASRPYNIYPHGITVRLPLYSRRLPKYKHLDPILFGEIFKRYWYTVDEGP 540
Db 481 LIIFKQASRPYNIYPHGITVRLPLYSRRLPKYKHLDPILFGEIFKRYWYTVDEGP 540
QY 541 TKSDRCITLRYYSFVNNERDLASGLIPGLICRESVDQGNQIMSKRVILFVSVD 600
Db 541 TKSDRCITLRYYSFVNNERDLASGLIPGLICRESVDQGNQIMSKRVILFVSVD 600
QY 601 NRSWYLTENIQFLPAPAGVLEDEPEQASNMHSINGVYEDSLQSVCLHEVAYWYLLS 660
Db 601 NRSWYLTENIQFLPAPAGVLEDEPEQASNMHSINGVYEDSLQSVCLHEVAYWYLLS 660
QY 661 IGAQTFELSVFSGYTFEYKHKWYEDTLFLPPSGTVMSENPGLMILGCHNSDFRNG 720
Db 661 IGAQTFELSVFSGYTFEYKHKWYEDTLFLPPSGTVMSENPGLMILGCHNSDFRNG 720
QY 721 MTALLKVSCKNTGDEYEDSYEDISAVILSKNNAIEPRSF 762
Db 721 MTALLKVSCKNTGDEYEDSYEDISAVILSKNNAIEPRSF 762
QY 762 ----- 762
Db 762 ----- 762
QY 781 PENDIEKTPWFARHRTPMFKIONVSSDLMLLRSPPHGLSLSDLOBAKYEPSPDDPS 840
Db 781 PENDIEKTPWFARHRTPMFKIONVSSDLMLLRSPPHGLSLSDLOBAKYEPSPDDPS 840
QY 762 ----- 762
Db 762 ----- 762
QY 841 PGALDSNNLSLSEMTFRPOLHSGDMVETPESGQLRLNEKIGTTAAETLKLDFRVSST 900
Db 841 PGALDSNNLSLSEMTFRPOLHSGDMVETPESGQLRLNEKIGTTAAETLKLDFRVSST 900
QY 762 ----- 762
Db 762 ----- 762
QY 901 SNNLSTIRSDUAGACTNTSSLSGPPSMVHYDSQDLDTTLFKKSSPLTESGPISLSE 960
Db 901 SNNLSTIRSDUAGACTNTSSLSGPPSMVHYDSQDLDTTLFKKSSPLTESGPISLSE 960
QY 762 ----- 762
Db 762 ----- 762
QY 961 NNDKSLLEGLNMQSSWGKNVSVSTESGRLFKGRRAHPALLTRDNALFKVYSILKNTN 1020
Db 961 NNDKSLLEGLNMQSSWGKNVSVSTESGRLFKGRRAHPALLTRDNALFKVYSILKNTN 1020

[illegible]

QY	1207	HEIKTGARQKRFSSLYIQFILMYSLDGKKNQYTRGNSGTGLMFFGNGVSSGJKNHIFN	1266
Db	2101	HEIKTGARQKRFSSLYIQFILMYSLDGKKNQYTRGNSGTGLMFFGNGVSSGJKNHIFN	2160
QY	1267	PEIIRYRLRLEHTHSISSTLMEJMGCDLNSCSMPLEKESKATSDAOITASSYFTNMFA	1326
Db	2161	PEIIRYRLRLEHTHSISSTLMEJMGCDLNSCSMPLEKESKATSDAOITASSYFTNMFA	2220
QY	1327	TWSPSKARLHLHGSRNAMPQVNNPKEMLAQYDFQTKMKVTGVTIQYKSLTISMYKFEFL	1386
Db	2221	TWSPSKARLHLHGSRNAMPQVNNPKEMLAQYDFQTKMKVTGVTIQYKSLTISMYKFEFL	2280
QY	1387	ISSSDQGHQWTLFPGNGYKVKYFQGNQDSFTFYVNSLDPPLLTRLRIRHPQSWHQIALRM	1446
Db	2281	ISSSDQGHQWTLFPGNGYKVKYFQGNQDSFTFYVNSLDPPLLTRLRIRHPQSWHQIALRM	2340
QY	1447	EVLGCEADPLY 1457	1506
Db	2341	EVLGCEADPLY 2351	2400
RESULT	2		
A47004	coagulation factor VIII precursor - mouse		
C:Species:	Mus musculus (house mouse)		
C:Date:	09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 31-Oct-1997		
C:Accession:	A47004		
R:Ellder, B.;	Lakich, D.; Gitschler, J.		
Genomics 16,	374-379, 1993		
A:Title:	Sequence of the murine factor VIII cDNA.		
A:Reference number:	A47004; MUID:93300511		
A:Accession:	A47004		
A:Status:	preliminary		
A:Molecule type:	mRNA		
A:Residues:	1-2319 <EUD>		
A:Cross-references:	GB:I05573; NID:g192456; PID:g192457		
C:Superfamily:	coagulation factor VIII; discoidin I amino-terminal homology; ferroxi-		
F:1-19/Domain:	signal sequence #status predicted <SIG>		
F:23-349/Domain:	ferroxidase repeat homology <FO1>		
F:402-730/Domain:	ferroxidase repeat homology <FO2>		
F:1686-2006/Domain:	ferroxidase repeat homology <FO3>		
F:2007-2156/Domain:	discoidin I amino-terminal homology <DN1>		
F:2160-2313/Domain:	discoidin I amino-terminal homology <DN2>		
Query Match	80.3%; Score 6261; DB 2; Length 2319;		
Best Local Similarity	53.5%; Pred. No. 0;		
Matches 1244:	Conservative 97; Mismatches 109; Indels 876; Gaps 9;		
QY	1	MOIEISTCEFFCLLRFCSATRRYYLGAVELSMDIYQSD-IGELPEVDARPPRPVKSEPF	59
Db	1	MOIALFACFEFLSLFNFCSSAIRRYLGAVELSWNIQSDLSVLHTDSRFLPRMSTSEPF	60
QY	60	NNSVYKKTLEFEPDHLFNIAKPRPMGGLGPIQAEVYDVTYITLKNNASHPVSLHA	119
Db	61	NNSIMKTVEYETEDQLEFNIAKPRPMGGLGPIQAEVYDVTYITLKNNASHPVSLHA	120
QY	120	VGVSYWKASEGAEDYDQTSOREKEDDKVFPGGSHYYVQVLEKNGPMASDPLCLTYSLS	179
Db	121	VGVSYWKASEGEDEYEDQTSOREKEDDKVFPGBSHYYVQVLEKNGPMASDPLCLTYSLS	180
QY	180	HYDLYKDLNSGLIGALLVCRESGLAKKEQTQLHAKITLLFAYFDEKSKSHSETKNSLMDR	239
Db	181	HYDLYKDLNSGLIGALLVCRESGLAKKEQTQLHAKITLLFAYFDEKSKSHSETKNSLMDR	240
QY	240	DASARAMPKMTVNGVYNRSLPGLIGCHRSVYHNVGMGTTPVHSIFLEGHTFLYRN	299
Db	241	DASARAMPKMTVNGVYNRSLPGLIGCHRSVYHNVGMGTTPVHSIFLEGHTFLYRN	300
QY	300	HROASLEISPIITFLTAQTLMDLGOFLLECHISSHQHDGMEAYVQVSDCPPEPOLRMK-N	358
Db	301	HROASLEISPIITFLTAQTLMDLGOFLLECHISSHQHDGMEAYVQVSDCPPEPOLRMK-N	360

QY 339 NEAADYDDDLTSDMDVVRDDDNSSPFIQIRSAVAKKPKTWVYIAAEEEDMDYAPLV 418
 Db 361 NEEMDYDDDLTSDMDVVRDDDNSSPFIQIRSAVAKKPKTWVYIAAEEEDMDYAPLV 418
 QY 419 LAPDRSKSOYLNNQPPORIGRRKRVKVFMAATDEFTKTRALIOHESGILGPLYGEVD 478
 Db 419 PISDQSGYSQYLNNQPPORIGRRKRVKVFMAATDEFTKTRALIOHESGILGPLYGEVD 478
 QY 479 TLLIIFKNQASPPYNIYPRGIDTVAPLYSRRLPKGVKHLKDPILPGLFFKTKWVYED 538
 Db 479 TLLIIFKNQASPPYNIYPRGIDTVAPLYSRRLPKGVKHLKDPILPGLFFKTKWVYED 538
 QY 539 GPTKSPDRCLTYSSSEFVYMERDLASGLIPGLICKESEVDORGNQMSDKRNVLLFSVF 598
 Db 539 GPTKSPDRCLTYSSSEFVYMERDLASGLIPGLICKESEVDORGNQMSDKRNVLLFSVF 598
 QY 599 DENRSWYLTENIORLPLNPAGVQLEDPFOASNMHSINGVYFDSLOLVCLEHVAWYI 658
 Db 599 DENRSWYLTENIORLPLNPAGVQLEDPFOASNMHSINGVYFDSLOLVCLEHVAWYI 658
 QY 659 LSTGAOTDLSVFFSGYTTKHKMYEDTTLTLPFSGEYTFMSNENGLMILGCHNSDFRN 718
 Db 659 LSTGAOTDLSVFFSGYTTKHKMYEDTTLTLPFSGEYTFMSNENGLMILGCHNSDFRN 718
 QY 719 RGMATALLKVSODKNTGDEYEDSISAYLLSKNNAIEPRSPSON----- 765
 Db 719 RGMATALLKVSODKNTGDEYEDSISAYLLSKNNAIEPRSPSON----- 765
 QY 779 TTPKNDMEKLEPQEEIAEMLKQSVSVSDMLMLLGQSHPTPHGLFLSDGOEAIYEAID 838
 Db 779 TTPKNDMEKLEPQEEIAEMLKQSVSVSDMLMLLGQSHPTPHGLFLSDGOEAIYEAID 838
 QY 765 ----- 765
 Db 839 DHSFNAIDSNEGSKYVQLRPESHSEKIVFTPOGLQLRSNKSLETTIEVKKKIGLOV 898
 QY 765 ----- 765
 Db 899 SLSPLNMTTLLSDNLKATFEKTDSSGPPDMFVHSSSKLSTAFGKAVSLVSHVPLN 958
 QY 765 ----- 765
 Db 959 ASEBDSNILDSTLMYSQESLPRDNLISLENDRLREKRFHGIALLTKDNTLFKDNVSL 1018
 QY 765 ----- 765
 Db 1019 MKTKNTYNHSTNEKLTESPTSTENSTDLQDALIKVNSEIQEVALIHDTGLGKNSI 1078
 QY 765 ----- 765
 Db 1079 YLRNLHMLNRTSTKKNKDIHRKDEDPIDODEENTIMPFKMLFESSNMWFKKTNGNS 1138
 QY 765 ----- 765
 Db 1139 LNSEQHSPPQOLVYLMFKYKKNOSFLSEKNKTVEDGFTKINGLKDMAFPHNMSTFLT 1198
 QY 765 ----- 765
 Db 1199 TLSVHENGHRNOKNIOEIELEKALIEEKVLPQVHEATGSKNFLKDILILGTRQNSL 1258
 QY 765 ----- 765
 Db 1259 YEVHVPYLQNTSINNSTNTVOIHMEHFRRKDKETNSEGLVNTREMYKNYPSOKNIT 1318
 QY 765 ----- 765
 Db 1319 TQSKRALGQRLSTQWLKINCSTQCLIKQIDSKEMKKTITSSLSDSVIAKSTQTN 1378
 QY 765 ----- 765
 Db 1379 SSDSHVKTSAFPPIDLKSPFQNKFSHVQASVITYDEKTSRSRIQESNNFLAKETKINP 1438
 QY 765 ----- 765

Db 1439 SLALPWNMFIDQKFTSPGKSNVNTSVYAKKRENIIFLKPILPEESGKIELLPQVSIQOE 1498
 QY 765 ----- 765
 Db 1499 EILPETSCHSPGHLNLMKEVFLQKIQGPIKWKAKRHGESIKGTESKNTSKLLNH 1558
 QY 765 ----- 765
 Db 1559 AMOVHAAQIPKDWKSKESPELISIKQEDTILSLPHGNSHIGANEKOMPREETW 1618
 QY 765 ----- 765
 Db 1619 VKOGOTORTSQIPYVLRKQOREL-SAFQSEQEDADYDALTIE-TIEFDLYSEDIKQ 1675
 QY 812 SPFSFOKTRHYFAAVERLMDGMSPPVLRNRAQSGSVPOFKKVFQEFIDGSETPOR 871
 Db 1676 GPRSFQOQTRHYFAAVERLMDGMSPPVLRNRAQSGSVPOFKKVFQEFIDGSETPOR 1734
 QY 872 LYGELNEHGLGLAPYTRAVEDNINAVTFRNQASRPYSFSSLSIEEDQGAEPKRN 931
 Db 1735 LYRGEINHEGLGLAPYTRAVEDNINAVTFRNQASRPYSFSSLSIEEDQGAEPKRN 931
 QY 932 VKPNETKYTFKVOHHAAPTKDEPCKAMAFSDVDLEKQVHSGILGPILVCHNTLNP 991
 Db 1794 VKPNETKYTFKVOHHAAPTKDEPCKAMAFSDVDLEKQVHSGILGPILVCHNTLNP 991
 QY 992 HGRQVVOEALFTIPDETQSWYTFTEMERNCRAPCNIQMEDPTEFKNYRFAHNGYIM 1051
 Db 1854 HGRQVVOEALFTIPDETQSWYTFTEMERNCRAPCNIQMEDPTEFKNYRFAHNGYIM 1051
 QY 1052 DTLPGVNAODORIRWYLLSGNSNENIHSIFSGHVTFRKKEEYKALNLYGVEYV 1111
 Db 1914 DTLPGVNAODORIRWYLLSGNSNENIHSIFSGHVTFRKKEEYKALNLYGVEYV 1111
 QY 1112 EMLPSKAGIMVECLIGELHLAGMSTLEFLVYSNKQCPPLGMAASHIRDFOTASQYQOW 1171
 Db 1974 EMLPSKAGIMVECLIGELHLAGMSTLEFLVYSNKQCPPLGMAASHIRDFOTASQYQOW 1171
 QY 1172 APLKARLHYSGSINAMSTKEPFSWIKVDLAPMTHGKITQGANOKFSSLYISOPTIWS 1231
 Db 2034 APLKARLHYSGSINAMSTKEPFSWIKVDLAPMTHGKITQGANOKFSSLYISOPTIWS 1231
 QY 1232 LDGKMWQYIRKGSSTGLWVFGNDVSSGIRKHNIFPILATYIRLHPHYSIRSLRMEI 1291
 Db 2094 LDGKMWQYIRKGSSTGLWVFGNDVSSGIRKHNIFPILATYIRLHPHYSIRSLRMEI 1291
 QY 1292 MGCIDLNSCSPILGMSKYSIDQIYASSYFTNMFATWPSQARHLQGTNAMRQVNDP 2213
 Db 2154 MGCIDLNSCSPILGMSKYSIDQIYASSYFTNMFATWPSQARHLQGTNAMRQVNDP 2213
 QY 1352 KEMLOYDFQKTKMYGVTYQYKSLTSMYKAEFLISSQDGHQWTLFPQNGKRVYFQGN 1411
 Db 2214 KEMLOYDFQKTKMYGVTYQYKSLTSMYKAEFLISSQDGHQWTLFPQNGKRVYFQGN 1411
 QY 1412 QDSFTPVNSLDPPLTRYLIRHPOSWVHQAIALRMEVGCCEADQY 1457
 Db 2274 QDSFTPVNSLDPPLTRYLIRHPOSWVHQAIALRMEVGCCEADQY 1457

RESULT 3
 KRU5
 coagulation factor V precursor - human
 N:Alternate names: coagulation labile factor; proaccelerin
 C:Species: Homo sapiens (man)
 C:Date: 19-May-1989 #sequence_revision 02-Jun-1995 #text_change 31-Oct-1997
 C:Accession: A56172; A42344; A28028; A27498; A25897
 R:Clone, L.D.: Moore, K.D.; Kane, W.H.
 Biochemistry 31, 3777-3785, 1992
 A:Title: Structure of the gene for human coagulation factor V.
 A:Reference number: A42344; M01D:9223268
 A:Accession: A56172
 A:Molecule type: DNA

A:Residues: 1 2224 <CR1>
A:Cross-references: GB:J05368
A:Accession: A42344
A:Molecule type: DNA
A:Residues: 48-58/79-89;120-130;191-201;239-249;313-323;368-378;428-437;461-471;533-542;
207070.2111-2120;2172-2181 <CR2>
R:Jenny, R.U.; Pittman, D.D.; Tooie, J.U.; Kriz, R.W.; Aldape, R.A.; Hewick, R.M.; Kauf
Proc. Natl. Acad. Sci. U.S.A. 84, 4846-4850, 1987
A:Title: Complete cDNA and derived amino acid sequence of human factor V.
A:Reference number: A28028; MUID:87260886
A:Accession: A28028
A:Molecule type: mRNA
A:Residues: 1-837,'R',859-864,'R',866-924,'E',926-1763,'I',1765-2212,'T',2214-2224 <JEN>
A:Cross-references: GB:M16967
A:Note: parts of this sequence, including the amino end of the mature protein, were dete
R:Kane, W.H.; Ichinose, A.; Hagen, F.S.; Davie, E.W.
Biochemistry 26, 6508-6514, 1987
A:Title: Cloning of cDNAs coding for the heavy chain region and connecting region of hum
A:Reference number: A27498; MUID:88107560
A:Accession: A27498
A:Molecule type: mRNA
A:Residues: 1-1284,'I',1286-1600 <KAN>
A:Cross-references: GB:M17785
A:Note: parts of this sequence were determined by protein sequencing
R:Kane, W.H.; Davie, E.W.
Proc. Natl. Acad. Sci. U.S.A. 83, 6800-6804, 1986
A:Title: Cloning of a cDNA coding for human factor V, a blood coagulation factor homolog
A:Reference number: A25897; MUID:86313665
A:Accession: A25897
A:Molecule type: mRNA
A:Residues: 1188-1215,1315-2224 <KA2>
A:Cross-references: GB:M14535
A:Note: parts of this sequence were determined by protein sequencing
R:Keller, F.G.; Ortel, T.L.; Quinn-Alten, M.A.; Kane, W.H.
Biochemistry 34, 4118-4124, 1995
A:Title: Thrombin-catalyzed activation of recombinant human factor V.
A:Reference number: A56139; MUID:95210278
A:Contents: annotation; thrombin cleavage sites
C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C:Genetics:
A:Gene: GDB:F5
A:Cross-references: GDB:119896; OMIM:227400
A:Map position: 1q23-1q23
A:Intons: 53/2; 84/1; 125/1; 196/1; 244/1; 318/1; 373/2; 432/3; 466/1; 537/3; 588/1; 65
C:Function:
A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa prot
A:Pathway: blood coagulation
C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma;
1-28/Domain: signal sequence #status predicted <SIG>
29-2224/Product: coagulation factor V #status predicted <MAN>
29-737/Product: coagulation factor Va heavy chain #status experimental <VAH>
F:29-737/Domain: A1 <DA1>
F:29-737/Domain: ferroxidase repeat homology <FO1>
F:33-329/Domain: ferroxidase repeat homology <FO2>
F:346-691/Domain: A2 <DA2>
F:351-684/Domain: ferroxidase repeat homology <FO2>
F:692-1573/Domain: B <DOB>
F:1183-1461/Region: 9-residue repeats (O-X-T/N-L-S-P-D-L-S)
F:1574-2224/Product: coagulation factor Va light chain #status experimental <VAL>
F:1574-1905/Domain: A3 <DA3>
F:1581-1905/Domain: ferroxidase repeat homology <FO3>
F:1667-1765/Region: phospholipid binding #status predicted
F:1906-2064/Domain: C1 <DC1>
F:1906-2061/Domain: discoidin I amino-terminal homology <DN1>
F:2065-2224/Domain: C2 <DC2>
F:2065-2221/Domain: discoidin I amino-terminal homology <DN2>
F:517.55.239.297.460.468.554.741.752.760.776.782.821.938.977.1074.1083.1103.1106.1479.149
F:167.193.248.329.500.526.603.684.1725-1751.1907-2061.2066-2221/disulfide bonds: #status
F:334-335/Cleavage site: Arg-Asn (protein C) #status predicted
F:363.693.1546/binding site: sulfate (Tyr) (covalent) #status predicted
F:376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
F:382.1338/binding site: carboxydrate (Asn) (covalent) #status absent
F:534-535/Cleavage site: Arg-Gly (protein C) #status predicted

```

F:1737-738/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F:15046-1047/Cleavage site: Arg-Thr (coagulation factor Xa, thrombin) #status experimental
F:15046-1047/Cleavage site: Arg-Ser (thrombin) #status experimental

```

Query Match	30.7%	Score 2391.5	DB 1	Length 2224
Best Local Similarity	26.1%	Fred. No. 1.3e-155		
Matches 593	Conservative 277	Mismatches 485	Indels 923	Gaps 35
QY	22 RRYVGAVELSDVYQOSDLGELPVDAREPPRPKSGFPNTSVV-YKTLVEFTLHLENI	80		
DB	32 RQFYVAAGISMSYRPE-----PTMSLNLSTYSRKKIYRREYR-ERK	75		
QY	81 AKRPPRMGLGPTIOAEYDVTVTLLKNMASHVPSLAVGVYKWSAGAEYDDQTSOR	140		
DB	76 EKRPQSTISGLGLPTLYAEVGDIIRKHFKNKADKPLSHPOGIRYSKLSGASYLDHTEPA	135		
QY	141 EKEDKVPGGSHYVWQVLYKENGMSADPLCLTYSYSHDLYVMDLMSGLIGALLVRE	200		
DB	136 EKMDADVAPGRKRYTTEWSTSEDGCTHDDPCLITITYSHENLIDFNSGLGLPLICK	195		
QY	201 GSIAEKTO-TLAK-FILLFAVEDEKSMHSEKTNLSLQDRDASAARAPKMHYNGYVN	258		
DB	196 GYLTEGCGQKTEFDKQDYLVLFAFVDESKSMQSSS-----LMTYVNGYVN	239		
QY	225 RSLPLGLCKRKSYYWHYVIGMTTEVHSIFLEGTEFLYRNHRQASLEISPTFLTAOTL	318		
DB	240 GMPDITVCANDHISWHLLGMSGPELFSIHENGQVLEQNHKVAISATLYASTTANMT	299		
QY	319 LMDLQPLFLCHISSHQDMEAYKVQVSCPEEPLRMKNNEADYDDDLIDSEMDYVR	378		
DB	300 VPEEGKWIISLTPEPHLOAQMOATIDINKCPKTKTNLKITFE-----	343		
QY	379 FDDDSPSFIQIRSVAKKHPKTVWYIAAEEEDMDYAPVLAPDDRYSKQYLANGPORI	438		
DB	343 -----QRRHKRMEYFIAAEVITWYAPVPIRANMDKRYRSQHLDFNSQI	387		
QY	433 GRKKYKVPFMATDTEFTKREAL---QHESGLGLGLGEVDDTLILFKNOASRPNYI	495		
DB	388 GHHYKVMYTOYEDNSF-TKHTVNPNMKEDGLGPIIRAOVADTLIKYFKNNASRPSTLY	446		
QY	496 PGGIT-----DVRELYSRLLPKGYKHKDPLILGELFKYKWTYVEDGPKRSDPRCLT	549		
DB	447 PGGVTFSPYEDVNSSTF---SGNNMIMIRAVOGCEYITTYKMNLLFEDPEPENAQCIT	502		
QY	550 RYSSVFANMERDLASGLIGPLILICYKESVDQRQNOIMSDKRNVLFSYFENRSRYLTEN	609		
DB	503 RRYVDYDIDMRDIAAGLIGLILICKSRSLDRRGIORADIEQAAFAFAVEDENKSWYLEDN	562		
QY	610 IQRFLPNPAGVQLEDEPQASVIMHSINGYVDSL-QLSVCLAEVAYWYILSIGAQDTFL	668		
DB	563 IAKFENDEPVARDPKPKFESNMIMSTINGYVESITTLGFCFDDTVQWHPFCVGTONEIL	622		
QY	669 SVFFSGYTFKHAMVYEDTLTFPFGSETYFMSMENPGLIILGCHNS-----	715		
DB	623 TILHFGHSIYKRRHEDITLTFPMNGESVITYMDVGMVLMILSMNSSPRSKRLKLFQVY	682		
QY	715 -----DFNRR-----	720		
DB	683 KCIIPDDEDEYEIEPPESTVATRKMDRLPEDEEESDADYQONRLAALGIRSFNS	742		
QY	720 -----GMAL-----	725		
DB	743 SLNQEEBEFNLTALLLENCTEFVSSNTDITIGVSNYSSPSNISKFTVNNLAEFOKAPSHQO	802		
QY	725 -----LKVSCDKNTGDIYEDSYED-----	745		
DB	803 ATTAGSPRLHLGLKNSVYLNSSTAHSSSPYSEDPIDPLQDPVDTGIRLLSLGAEKFSQEH	862		
QY	745 -----	745		
DB	863 AKHKGPYERDQAAKHREFSWMLLAHKVQRLHSQDTGSPSGMRPVEDLPQDGTGSPSRMR	922		

QY 745 -----ISAFLSKNNA----- 756
Db 923 PKKDPSSLILLKQSNSSKILVGRWHLASEKSGYELIIOPTDEDTAVNMVLISPONASRAW 982
QY 756 ----- 756
Db 983 GESTPLANKPGKSGHPKPRVRRHKSLOYRQDQSGKSLKKSQFLIKTRKKKEKHTHAP 1042
QY 756 IPRSF-----SQ 763
Db 1043 LSPRTFHLRSEAYNTFFSERRLKSLYLAKSNETSLPTDNLQTLPSMDGWLASLPDHQ 1102
QY 764 N----- 765
Db 1103 NSSNDTGQASCPGLYQIVPPEHHYQTFPIQDDQMHSTSDPSHRSSPELSEMLEYDRS 1162
QY 765 -----PVLKRA- 772
Db 1163 HKSFTDLSQMSPSSEHEWQTVISPDLISQVTLSPELISQTLNLSPLDLSHTTSLPELIQRNL 1222
QY 772 ----- 772
Db 1223 SPALGOMPISPLDLSHTTSLDLSQTLNLSPLDLSQTLNLSPLDLSH 1282
QY 772 ----- 772
Db 1283 TLLSLDFQSTNLSPELSHMTLSPELSQTLNLSPALGOMPISPLDLSHTTSLDLSQTLNLSPE 1342
QY 772 -----OREITRTTLOSQOEI----- 788
Db 1343 LSQTNLSPALGOMPISPLDLSHTTSLDLSQTLNLSPELSQTLNLSPLDLSHMTLSQTLNLSPLDLS 1402
QY 788 ----- 788
Db 1403 TPDLQMTLSPLDLSHTTSLDLSQTLNLSPELSQTLNLSPLDLSHTTSLDLSQTLNLSPLDLS 1462
QY 788 -----DYDT 792
Db 1463 ITPPSSSOSLLLOEPNESFPYDGLGOMPSPSSPILINDTFLSKENPLIVGLSKDGTXY 1522
QY 793 ISEVKKKE-----DFDIYDE-----DENOSP-----RSFOKTRHY 823
Db 1523 IETIKKEVQSSDDYAEIDYDYPDPTDYRTINNSRDNDNLAWLKRSNNNGRRY 1582
QY 824 FLAAYERLMDYMSSPHYLNRN--AQSGSPQ--FKVVFQETDGSFTQPLRGELN 878
Db 1583 YIAAEISMDY-----SEVQRETDIEDSDIDEDTYYKRVFRKYLDSFTKRDGEYE 1638
QY 879 EHLGLGPIRAVEEDNIMVTFRNOASRPYSFYSLSIE-----EDQOGAEPKRN 930
Db 1639 EHLGLGPIRAVEEDNIMVTFRNOASRPYSFYSLSIE-----EDQOGAEPKRN 930
QY 931 FVKPNETKTYFKKVOHHAAPTRDEFDCKAMAYFSDVLEKDVHSGILGILVCHTNTLN 990
Db 1699 AVQPSSTTYVHAHERSPESPSGACRAMAYSAVNPRKDDHSGILGILVCHTNTLN 1758
QY 991 AHGRVYTOEFLFTTIDETKSMFTFEMERNCAPCNIQMEDPTFKENRYFAINGYI 1050
Db 1759 DSNMPLDMNEFVLFTMTPEKKSUYEKKRSRWSR-----LTSSEMKSHFHAINGMI 1812
QY 1051 MDLTLGLVAAOQRIRWYLLSGNSNENHSHFSGHVTFYRKKEVKMALVLYLGVFET 1110
Db 1813 YSLPGLKMYEDEMWRKHLNLNGSSODIHVHFHGOITLNGKNKHQDLQVWLPGLGSFET 1871
QY 1111 VEMLPKAGIWEVECLIGELHGLHASTLELVYNNKQCPFLGASGHIRFOITAGSQQYGO 1170
Db 1872 LEMKSKRGWMLNLTEVGNQARAGMOTPLINDROCRMNGSLGILSISOQKASEFLGY 1931
QY 1171 WAPKLARLHSGSINAMSTKE--PFS--WIKVDLAPMIHIGIKTOGAROKFSFLYIS 1224
Db 1932 WEPRLARLNNGSYANWSEKILAEAFASKPMYQIOWMOKREVIITIGIOQAKRYKLSCYT 1991
QY 1225 QFIWYSLDGKKWQYIRGNSGTCLLWVFGONVDSGKHNHFNPIIARIIRLHPHYISIR 1284

Db 1992 ERYAYSSNQINMOIFKGNSTNRWYFNGNSDASTIKENQDPPIVAYYIRISPRAYNR 2051
QY 1285 STIRBELMGCDLNSCMPLGHEKRAISDAQITASAYFNMAF--WSPKARLHLOGRNA 1343
Db 2052 PLRLLELOGCEVNGSTPLGHEKRIENKQITASSFKMSGWDYWEPRARLNOGRNA 2111
QY 1344 WRPQVNNPKEWLQVDEQTKMYVTGVTTOGVKSLTSMYKEFLISSQDGHQWTLFPONG 1403
Db 2112 WQAKNNNNKQWLEIDLTKIKKITAITITOGKSLSEMYKSYTHIYSPQWEMKPYRLKS 2171
QY 1404 KY--KVFGNDSFTPPVYNSDPLTLRYLRIHQSWHQAOLNMEVYGCACADLY 1457
Db 2172 SWDKIFEGNTNTGKHNKFNPNPILSRIRVIRPTWQSIALLRLEFGC---DIY 2224
RESULT 4
KRB05
coagulation factor V precursor - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Mar-1993 #sequence, revision 28-Apr-1995 #text, change 20-Mar-1998
C:Accession: A42580; A36497
R:Guinto, E.R.; Esmon, C.T.; Mann, K.G.; Macgillivray, R.T.
J. Biol. Chem. 267, 2971-2978, 1992
A:Title: The complete cDNA sequence of bovine coagulation factor V.
A:Reference number: A42580; MUID:92147638
A:Accession: A42580
A:Molecule type: mRNA
A:Residues: 1-2211 <GUI>
A:Cross-references: GB:M81440; NID:g163037; PID:g163038
A:Note: sequence extracted from NCBI backbone (NCBIN:80774, NCBI:P:80776)
R:Katafatis, M.; Jenny, R.J.; Mann, K.G.
J. Biol. Chem. 265, 21580-21589, 1990
A:Title: Identification and characterization of a phospholipid-binding site of bovine
A:Reference number: A36497; MUID:91072354
A:Accession: A36497
A:Molecule type: Protein
A:Residues: 1566-1570, 'X', 1572-1581, 'X', 1583-1584, 1673-1676, 'X', 1678-1679, 'X', 1681, 'X'
R:Xue, J.; Katafatis, M.; Silveira, J.R.; Kung, C.; Mann, K.G.
Biochemistry 33, 13109-13116, 1994
A:Title: Determination of the disulfide bridges in factor Va heavy chain.
A:Reference number: A55979; MUID:95034740
A:Contents: annotation
A:Note: 566-Cys and 617-Cys were shown to have free sulfhydryls
A:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C:Function:
A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa p
A:Pathway: blood coagulation
C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasm
F:1-88/Domain: signal sequence #status predicted <SIG>
F:29-2211/Product: coagulation factor V #status predicted <MAT>
F:29-741/Product: coagulation factor Va heavy chain #status predicted <VAB>
F:29-345/Domain: A1 <DA1>
F:33-329/Domain: ferroxidase repeat homology <FO1>
F:346-685/Domain: A2 <DA2>
F:351-688/Domain: ferroxidase repeat homology <FO2>
F:696-1564/Domain: B <DOB>
F:1175-1437/Region: 9-residue repeats (O-X-T-N-L-S-P-D-L-S)
F:1565-2211/Product: coagulation factor Va light chain #status predicted <VAL>
F:1572-1892/Domain: ferroxidase repeat homology <FO3>
F:1654-1752/Region: phospholipid binding #status predicted
F:1893-2051/Domain: discoidin I amino-terminal homology <DN1>
F:2052-2211/Domain: C2 <DC2>
F:2052-2208/Domain: discoidin I amino-terminal homology <DN2>
F:167-193, 248-329, 499-525/Disulfide Bonds: #status experimental
F:226, 229, 297, 382, 460, 553, 587, 745, 756, 774, 780, 902, 952, 964, 1044, 1053, 1062, 1071, 1078, 110
F:334-335/cleavage site: Arg-Asn (protein C) #status predicted
F:363, 697, 1537/Binding site: sulfate (TYR) (covalent) #status predicted
F:376-377/cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
F:533-534/cleavage site: Arg-Gly (protein C) #status predicted

F:607-688,1712-1738,1894-2048,2053-2208/Disulfide bonds: #status predicted
F:741-742/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predicted
F:1034-1035/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predicted
F:1564-1565/Cleavage site: Arg-Ser (thrombin) #status experimental

Query Match 30.4%; Score 2370; DB 1; Length 2211;
Best Local Similarity 26.2%; Pred. No. 3 9e-154;
Matches 594; Conservative 265; Mismatches 487; Indels 924; Gaps 35;

QY 22 RRYLAGVELSMDYMOSELGELPYDARPPRPVPSFPNTSVYKTLFVEFDHLENIA 81
D 32 ROYVAAQSIKRWNR-----PESTHL-----SSKPEFS--FKKLYREY-EAYQKE 76
QY 82 KPRPPMGLLGPITQAEVYDTVTTLKNMASHPYSLHNAVSTWKASGEAEYDDQSQRE 141
D 77 KPSRTSGLLGPLYAEVGMKMYHFKNAKAPLSIHAOGIKYSKSEGAASYSDHTLPM 136
QY 142 KEDKVPGGSHTYVMQVLEKNGPMASDPLCLITYSLSHDYKDLNGLALNLCREG 201
D 137 KMDVAPGGEYTYEWLISHSGPTHDPPCLTHIYSYVNLDEFNSGLIGLICKKG 196
QY 202 SLAKETQTL--HKFLLFAVDEGKSMHSETKNSLMQDRDASARAMPKMTYNGYVR 259
D 197 TLNEDGQKMEKOHVLMFAVDESKMNOTS-----LMYNGIYNG 240
QY 260 SLPLGLCHRRKSYVHWYIGGTPPEVHSIFLEGHTFLVRNHRQASLEISPTFLAQTLL 319
D 241 TMDPITCAHDHISMHLIGSSGPELSIFHNGQVLEQNHHKISALIVSATSTANMTV 300
QY 320 MDGGEFLFCHISHOHDMGEAYKVYKDSCEPEPOLBMKNEAEADYDDLDSDMDYVR 379
D 301 SPBRKWTASLIPRHFOAGQAYIDIKNCAKTRNPK-----LTRQ----- 344
QY 380 DDNSPFIQIRSAVAKKPKTWYHIAEEDMDAYLVLAPDDRYSKOYLNNGPORG 439
D 344 -----RRHIRWEFFIAEYIMDYAPILIRANMDKKRSLHLDNFSRIG 388
QY 440 RYKKVAFMAVTDTEFTK--EAIQHEGSLGPLLYEGVDTLLIIRKNASRPYNTYPH 497
D 389 KHKVKVYKQYODSEFKRLEDEPSEGDGILPIIRAQVDTLKIYKKNASRSYSYPH 448
QY 498 GII-----DYRPIYSRLPKGVKHLMDPILRGEIRKRYKTYVDEGPTKSDPRCL 548
D 449 GVTFSPYDNEVNSSSTSGSTMIRAVR-----PGETYTKNLLISEPENDAOCL 500
QY 549 TRYYSFVNNERDLASGLIGPLICYESYDQNGQIMSKRNYLPSVDENSTVLT 608
D 501 TRPYSNVDTLRDLASGLIGLLILICKRSIDRKGIOAADIEQAVFAVEDENKSWITIED 560
QY 609 NIOFLNPAGVOLDPEFOASNIMH-----SINGYVDSIQ--LSVCLHEVAYWYIISIG 662
D 561 NIKFCENPEKVRKDRDPKYESINIMSNFTLPAINGYVESIPILGFCFEDDTVQMHFCSVG 620
QY 663 AQDFLSVFSGTYFKKMYEDTLTFPFSGETVFMSEMPGLMITIGCHNSDRANGMT 722
D 621 TQNDILIHGHSTIGKHEHDTLLFPMQGESVYTMNVGMULTMNASNSRSLKR 680
QY 723 ALKVSQCDKNTGD-----YEDS-----YEDISAVYLL-- 751
D 681 LRFADACINDDDDSEIITYEPBSGTAMTKTKIHDSSEIENDADSDIODELALILGL 740
QY 751 -----SKNNAIEPRFSQ----- 764
D 741 RSFNSSINQEKDELNTALALEKDESEFIPPSANRSLDSSSRSHVRLAKNFAESLK 800
QY 764 -----NPPVLK----- 770
D 801 TLHLLEPAAGSPLEHAGLDKNSALNPPMAHSSPYEDPREDHPLSDVYGVSLPFGTG 860
QY 770 -----RHQR-----EITRTLLQSDQ-----EET----- 788
D 861 FKNRPAKHQHFQYGRGOAAKHKFSQTRFAHKRTRRLSDQNSSSSRMGPMEDLPDLLL 920

QY 788 ----- 788
D 921 LQKDPYKILNGEMHLYSEKSYETIIOANENKTVNKLPSNPQNDRTWGENIPFKNSHG 980
QY 788 ----- 788
D 981 KQSGHPTLVTRKPKLQDRDRNRSKLKESPLIRTRKKKEKPRAYHVPFSRSHPLR 1040
QY 788 -----DYD----- 792
D 1041 GEVNASFDRRHHSLILHASNETSLIDLQTFPSMNLASLPHDQTSNDTTSQT 1100
QY 792 -----TISVEMKEDPEIYED-----ENSP----- 814
D 1101 SSPDPYTPVSPPEHYQIFPIQSDPTSTAPSNRSPDPTSTAPSNRSPPTOPSOIP 1160
QY 814 ----- 814
D 1161 NYDLNRRAIPTVSQIFPSLELEVWOTANSLSQPSISPDLCOMALSPDGOESLSPDL 1220
QY 814 ----- 814
D 1221 GQTSLSPLDSQESLSPDLGQTALSPDPSQESLSPDLGQTALSPDPSQESLSPDLGQTALS 1280
QY 814 ----- 814
D 1281 PDGOESLSPDLGQTSLSPDLGQESLSPDLGQTALSPDPSQESLSPDLGQTALSPDPSQ 1340
QY 814 ----- 814
D 1341 SLSPDLGQTSLSPLDGOESLSPDLGQTALSPDPSQESLSPDLGQTSLSPLDGOESLSPDL 1400
QY 814 ----- 814
D 1401 GQTALSPDLDSQESLSPDLGQTSLSPDLSESLSPDLSQLDLKQTSPLDLNQTSHSESS 1460
QY 814 ----- 814
D 1461 QSLPLPEFGQTFPNADIGQMPSPDSTLNTTIPPEEFNPLVYVGLSRDDGDYIEILPRQ 1520
QY 814 -----RSQKTRHYFAVER 830
D 1521 KEESSEEDYGEFFVAYNDYQTLRTDINSSRNPDNIAAWYLRNSTGNKRYIIAAEEI 1580
QY 831 LMDYGNSSPHVLRNRAQSGS---VPO---FKRYVPOEFGDSFTQPLYGELNEHLGL 884
D 1581 SWDS-----KPYQSDVDYVPEDTYKKYVFRKAYIDSTFTKLDPGGEYEHGIL 1631
QY 885 GPYIRAEVEDINVTFRNQAISREYSFSSLSIYE-----EDORGAEPKRNKVPNE 936
D 1632 GPVIRAEVDVIOVFRKNLASRPSLSHAHGLSYEKSEKTEYEDDSEPKEDNAIQPNK 1691
QY 937 TKTYFMVQHHMAPTDEDFCKAMAYPSVDLEKDVHSGLIGPLVCHTTLNPAHROY 996
D 1692 TYTYVMATTRSPENBNGSACRAMAYSAVAPKDIHSGILGILLICRKQTLKETMPV 1751
QY 997 TYOEFALFTIPDETYSKYTEMNERCARPCNIQMEDPFKEKRYRHAINGYIMDTLP 1056
D 1752 DMREYVLLFMVFPDEKSKWYIDKPTRSMRASS-----EYKSHHEHALINGMAYLN-LPG 1804
QY 1057 LVNAQQRIRMYLLSMGNSNENHISHSFGHVTVRKKEEYKMALYNLPGVFETVENLPS 1116
D 1805 LRMYDEQMVVRLHNLNIGSSDIDHVFHFGOTLLENGTOHQDLGWPLPDSFXTLEMKAS 1864
QY 1117 KAGIWRPECLIGELHAGMSTFLVYSNKGQTPLGMASSHIIRPQITASQOYQOMAKLA 1176
D 1865 KPGMWLLDTEVGETQKRGKQTPPLIVDRCKMPGSLTGLIASQIQASFGWGEWEKLA 1924
QY 1177 RLHYSGSINAM-----STK-EPSMWIKVDLLAPMIIGITQGAAROKFSLSYISOFTIMY 1230
D 1925 RLNNGSGYNMIAIEKLSTEENPEFWIOVDMQKQEVLLGLTQGAKHHLKRYIYTTTEPCVAY 1984

OY 1231 SLDRKKQTYRGNSTGLTLMFPGVNDSSGJKNHIFNPPIIARTIRLHPHTYSIRSTLRNE 1290
|||:::|||:::|||:::|||:::|||:::|||:::|||:
Db 1985 SLDRKMRNRIFFGNSTRRNMVFFGNSDASTIKENQIDPPVAARTRISPTSYNKPALRLR 2044
|||:::|||:::|||:::|||:::|||:::|||:::|||:
OY 1291 LMGDLLNCSPILGMESKAISAQAQTASSPYTNMFAT-WSEPSRKALHLQGRSNAMRPQVN 1349
|||:::|||:::|||:::|||:::|||:::|||:::|||:
Db 2045 LQGCEVNGCSTPLMGESKIENKOITASSPFSKMVGWMEPEFLA RLNAAGRVANMAOKAN 2104
|||:::|||:::|||:::|||:::|||:::|||:::|||:
OY 1350 NPKEWLGVDFOKTKMYGVTTGYVKSLTSMTKYVEFLISSODQHWTLPFGNQKV--KV 1407
|-:-||-:-||-:-||-:-||-:-||-:-||-:-||-:-||
Db 2105 NNMQWLQIDLKIKKTFTIVTGCKSLSSEMKYVSXTIHSDQCTDMKPRESMSWDKT 2164
|||:::|||:::|||:::|||:::|||:::|||:::|||:
OY 1408 FOGNDSTFPVNSLDPLELTRYLRIHQSVWHQIALRMELIGEADOLY 1457
|||:::|||:::|||:::|||:::|||:::|||:::|||:
Db 2165 FEQNMMVRGHAKNFEPNPITSFRITIPRTWNOSTIA LRLEFGC---DMY 2211
|||:::|||:::|||:::|||:::|||:::|||:::|||:

SULT 5

Nitrooxidase (EC 1.16.3.1) precursor - human

N.Alternate names: ceruloplasmin
N.Contains: ferroxidase long form (CP-1); ferroxidase short form (CP-2)
C.Species: Homo sapiens (man)
C.Date: 31-Aug-1990 #sequence_revision 12-May-1995 #text_change 05-Sep-1997
C.Accession: A25443; A24165; A00524; I59067
R.Koschinsky, M.L.; Funk, W.D.; van Oost, B.A.; Macgillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 5086-5090, 1986

A>Title: Complete cDNA sequence of human preceruloplasmnin.

A.Reference number: A25443; MUID:86259737

A.Molecule type: mRNA
A.Accession: A25443

A.Residues: 1-1060,1065-1069 <ROS>

C.Cross-references: GB:M13699; NID:g180256; PID:g180256

R.Note: This is the short or CP-2 alternatively spliced form

R.Mercer, J.F.B.; Grimes, A
FEBS Lett. 203, 185-190, 1986

A>Title: Isolation of a human ceruloplasmin CDNA clone that includes the N-terminal lead

A.Reference number: A24165; MUID:86275241

A.Molecule type: DNA

A.Accession: A24165

A.Molecule type: mRNA
A.Residues: 1-40:549-599; 784-829;919-952 <MER>
R.Yang, F.; Friedrichs, W.E.; Cupples, R.L.; Bonifacio, M.J.; Sanford, J.A.; Horton, W.A
J. Biol. Chem. 265, 10780-10785, 1990

A>Title: Human ceruloplasmin tissue-specific expression of transcripts produced by alter

A.Reference number: A35450; MUID:90285218

A.accession: A35450

Molecule type: DNA
M.residues: 1007-1064 <YAN>

A.Cross-references: GB:J05506

A>Note: This is the long or Cp-1 alternatively spliced form

R.Rakahashi, N.; Ortel, T.L.; Putnam, F.W.
Proc. Natl. Acad. Sci. U.S.A. 81, 390-394, 1984

A>Title: Single-chain structure of human ceruloplasmin: the complete amino acid sequenc

A.Reference number: A00524; MUID:84119493

A.accession: A00524

A.Molecule type: protein
A.Residues: 20-1060,1065-1069 <TAk>
A>Note: 79-Gly and 449-Gly were also found
R.Yang, F.; Naylor, S.L.; Lum, J.B.; Cutshaw, S.; McCombs, J.L.; Haberhaus, K.H.; McGilliv

Proc. Natl. Acad. Sci. U.S.A. 83, 3257-3261, 1986

A>Title: Characterization, mapping, and expression of the human ceruloplasmin gene.

A.Reference number: I59067; MUID:86205876

A.accession: I59067

A.Status: translated from GB/EML/DDBJ

A.Molecule type: mRNA
A.Residues: 218-1069 <RES>

C.Cross-references: GB:M13536; NID:g180248; PID:g180249

C.Comment: Ferroxi-dase is a blue, plasma alpha2-glycoprotein binding 6-7 copper ions per C;Comment: In Wilson's disease the plasma levels of ferroxidase are diminished or undetect

C;Genetics: The three fragment chains are produced spontaneously during purification and

A:Gene: GDB:CP
A:CROSS-references: GDB:I19069; OMIM:I17700

#:Map position 3q23-3q25
 A:Introns: 1006/3; 1061/1
 C:Note: the list of introns is incomplete
 C:Description: catalyzes the oxidation of free iron(II) to iron(III) coupled with the
 A:Note: iron(III), but not iron(II), is the form bound and transported by transferrin
 A:Note: other possible functions are amine oxidase activity, copper transport and hom
 C:Superfamily: ferroxidase; ferroxidase repeat homology
 C:Keywords: acute phase; alternative splicing; copper; duplication; glycoprotein; oxidi
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-1069/Product: ferroxidase, long form #status predicted <MATL>
 F:20-1060/Product: ferroxidase, short form #status experimental <MATS>
 F:20-499/Product: ferroxidase 67k chain #status experimental <K67>
 F:23-357/Domain: ferroxidase repeat homology <FER1>
 F:373-718/Domain: ferroxidase repeat homology <FER>
 F:501-905/Product: ferroxidase 50K chain #status experimental <K50>
 F:733-1059/Domain: ferroxidase repeat homology <FER3>
 F:907-1065/Product: ferroxidase 19k chain #status experimental <K19>
 F:138,397,762/Binding site: carboxylate (Asn) (covalent) #status experimental
 F:174-200,216-357,534-560,637-118,874-900/Disulfide bonds: #status predicted
 F:227,588,926/Binding site: carboxylate (Asn) (covalent) #status absent
 F:295,338,343/Binding site: copper (His, Cys, His) (type 1) #status predicted
 F:358/Binding site: carboxylate (Asn) (covalent) (partial) #status experimental
 F:656,699/704,709/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted
 F:994,1040,1045,1050/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted

[illegible]

Db 559 ICKGSLHANGROKDVKEFLPTVEDENESELLNIRMFETAPDQVDEKEDFOESN 618
 QY 632 IMHSINQVDFDSLO-LSVCLHEVAWYILSIGAOTDPLVSFGYTKRKHVYEDTLTLE 690
 Db 619 KMHSMNGFMGNQGLTMCCKDSVWYILFSAGNADVHGIFSGNTYLMRERDITANLF 678
 QY 691 PFSGEYFMSMENGLMILGCHNSDFRNRGTALTLLKVSCKDNKGYYEDYEDISAVYL 750
 Db 679 PQSLTLHMPDTEGTENVELLTDHTTGKQKTYNQCRQS-----EDS----- 726
 QY 751 SKNNALPFRSFSQNPVYLKRHQREITRTTLOSDEEIDYDDTISVEKKEDFDIYEDEN 810
 Db 726 ----- 726
 QY 811 QSPRSFOKTRHRYIAVERIMDYGMSSP-----HYLRNAGSGV-----P 853
 Db 726 -----TFLEGETTYIAAVEWMD-----SPQREWEKELHQLQDNVSAFLDKGEFYIGS 777
 QY 854 QFKKVFQEFYDGSFTOPLYRGELEHGLGPLYRAVEEDNINWTFRNOASRPYSFSS 913
 Db 778 KYKVVYRQYTDSTFRYVERKAEHEHGLIGPOLHADVGVKAKIKRNATRTYSIHA- 837
 QY 914 LISTEEDROGAEPRKNEVK--DNETKYFKVQHMAPTKDEDCAMAYFSVDLEK 970
 Db 837 -----HGVTESSTVPTLPGETLTYWKIPERSGAGTEDSACIPMAYYSTVDQVK 887
 QY 971 DVHSGILGPLLYCHTNLNPAGROYVQVEALFTTFIDFKSWYTFENMRNCRAPNT 1030
 Db 888 DLVSGLGLPLVCRRPYLKFNPRKL--EFALLFLVFDENESWLDNITYSDDHEKV 945
 QY 1031 QMEDPTREKNEFHAINGYIMDTLPGLVMAODRIWYILMSNENSHISFSGYEVY 1090
 Db 946 NKDEDEFIESKMAINGRMGNIQGLTMHGVDEWYLMKMGNEIDIAHTEFHGHSFYQ 1005
 QY 1091 RKKEEYKALYNLYPGVETVEMLPKAGIWRVBCILGEHLHAGMSTFLVYSKRCQTP 1149
 Db 1006 KHRGVYSSVDYDIPPGTYQTLMEFPRPGIWLHCHVTDHHAEMETTYTLQNGEPR 1064
 RESULT 6
 A35210
 ferroxidase (EC 1.16.3.1) precursor - rat
 N:Alternate names: ceruloplasmin
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 13-Mar-1998
 C:Accession: A35210; A41753; A29564; S21692
 R:Fleming, R.E.; Gitlin, J.D.
 J. Biol. Chem. 265, 7701-7707, 1990
 A:Title: Primary structure of rat ceruloplasmin and analysis of tissue-specific gene exp
 Accession: A35210
 Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-1059 <FILE>
 A:Cross-references: GB:J05424
 R:Fleming, R.E.; Gitlin, J.D.
 J. Biol. Chem. 267, 479-486, 1992
 A:Title: Structural and functional analysis of the 5'-flanking region of the rat cerulop
 A:Reference number: A41753; MUID:92112697
 A:Accession: A41753
 A:Molecule type: DNA
 A:Residues: 1-48 <FLD>
 A>Note: the authors translated the codon GAA for residue 40 as Gly, GAA for residue 41 a
 R:Aldred, A.R.; Grimes, A.; Schreiber, G.; Mercer, J.F.B.
 J. Biol. Chem. 267, 2875-2878, 1992
 A:Title: Rat ceruloplasmin. Molecular cloning and gene expression in liver, choroid plex
 A:Accession: A29564; MUID:87137545
 A:Molecule type: mRNA
 A:Residues: 'NSG', '215-216', 'Y', '218', 'FAT', '222', 'F', '224-226', 'E', '228', 'LL', '231', 'D', '233-235', 'R'
 A:Experimental source: liver
 A>Note: the authors translated the codon GCG for residue 60 as Gly and GTG for residue 1
 R:Ryan, T.P.; Grover, T.A.; Aust, S.D.

Arch. Biochem. Biophys. 293, 1-8, 1992
 A:Title: Rat ceruloplasmin: resistance to proteolysis and kinetic comparison with hum
 A:Reference number: S21692; MUID:92117681
 A:Accession: S21692
 A:Molecule type: protein
 A:Residues: 20-29, 'Q', 902-910 <RYA>
 C:Superfamily: ferroxidase; ferroxidase repeat homology
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-1059/Product: ferroxidase #status predicted <KAT>
 F:23-356/Domain: ferroxidase repeat homology <FER1>
 F:372-712/Domain: ferroxidase repeat homology <FER2>
 F:727-1053/Domain: ferroxidase repeat homology <FER3>

Query Match 21.8%; Score 1697; DB 2; Length 1059;
 Best Local Similarity 32.5%; Pred. No. 2,1e-108;
 Matches 386; Conservative 188; Mismatches 428; Indels 186; Gaps 23;

QY 5 LSTCFPLCLRFCSARRRYILGAVELSWDMQ-SDGEL-PVDARPPRYKSPRNTS 62
 Db 6 LSAFLFL-HSSLAWTRKKHYIGITEAWDYASGESEKELISVDESNFYLRNGPRIG 64
 QY 63 VYKKTLEFERTDHLFNLAKPRPMGLIPTIOAEYVDVITLKMAHSPVGLHAGV 122
 Db 65 RYKKAIXSETDCTFTKTDKPAWLGFLGVIAEYGDKYSVAKNFASRPTFFHAGV 124
 QY 123 SYWASSGAEDDQTSQREKEDKVPFGSHYVQYLKENGPNASDPLCLTYLSHYD 182
 Db 125 TYTKANEGAIYDWTDFGRADKLFPQGYLYLRA-NEPSPEGDSNCTRIYHSHVD 183
 QY 183 LVKRLNGLIGALVCBEGSLAKETQTL-HKFTLLAVFDEGSMSETKNSLM----- 237
 Db 184 APKDIASGLIGPLILCKKSLHKEKENIDQEFVLMFSYVDENISWYLEDIKTFCSPE 243
 QY 237 -QORDASARAWPMHNVGVNRSPLGLGCHKRSYWHVIGCTPEVHSIFLEGTTF 295
 Db 244 KYVDKDNEDFOESNMYTSINGTTFGSLPGLSCADRYKATLEFGNGEYDVALFPGAL 303
 QY 296 LVNRHROASLEISPTLETLAQTLLMDLGOFLFCHISSHODGMAVYKVDSCPEQLR 355
 Db 304 TSKNYHTDIILFPALILDVSMVANOQGVWMLSCQNLNHLKAGLQAFQVQDC----- 357
 QY 356 MKNNEAEVDLDDLTDEMDVFRDDDNSSFIQIRSYAKKHPTWYHLYAAEEDMDYA 415
 Db 357 ---NKPSP---DDDIQDRHV-----RH-----YYIAEETIMDYA 385
 QY 416 P-----LVLPDDRYSKQYILNNGPQIRGKYKRYKREAYDETF---KTRERAI 461
 Db 386 PSCGDTTGENLISLGSDSRVEEQ---GATRIIGSIRKLVREYDDSTNRKRGPD 441
 QY 462 QHESGILGPLLYGEGDTLLIFKNOASRPYNIYPHGITDVR---PLYSRLPKGVKHL 517
 Db 442 EHHGILIGPVYMAVGVQIINTFHNKQFPLSTIOPMKVRFKMEGNYG---PDGSSK 498
 QY 518 KDPFLPELIFKYWTYVEDGPTKSPRCITRYSYFVNKNERDLASGLIGPLLYKES 577
 Db 499 QASHVAKEFTYETWYVPEKMGPRYADPVCLSKMYGVDLTKDIFGLIGPMKICKGS 558
 QY 578 VDGNGIMSDKRVNVLFSVDENRSWYLTENIORFLPNPAGVQLEDEPQASIMNSIN 637
 Db 559 LLDAGROKDVKEFYLRATVFDENESILLDNIMFTAPENVDKEDDEFOESNKMHSN 618
 QY 638 GYVDFDSLO-LSVCLHEVAWYILSIGAOTDPLVSFGYTKRKHVYEDTLTLEPFGSET 696
 Db 619 GFKNGLNPLGMKMGESVWTLFESAGNADVHGIFSGNTYLSGERDITANLFPHSLT 678
 QY 697 VMSMENPGLMILGCHNSDFRNRGTALTLLKVSCKDNKGYYEDYEDISAVYLSKNNAI 756
 Db 679 LLMPTDEGSDVDCLTLDHTGTGKQKTYNQCRQS-----KQGFEDVTLY----- 723
 QY 757 EPRFSQNPVYLKRHQREITRTTLOSDEEIDYDDTISVEKKEDFDIYEDENQSPRSF 816

Db 723 ----- 723
QY 817 OKTRHFIATAEVRLMDYGMSSSPHYLRNROSGSY-----POFKVVEOF 863
C:Accession: JC4915
B:Ogura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanai, Y.
Biochem. Biophys. Res. Commun. 225, 932-938, 1996
A:Title: Cloning and expression of cDNA for O-acetylation of GD3 ganglioside.
A:Reference number: JC4915; MUID:96374422
A:Accession: JC4915
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-427 <OGU>
A:Cross-references: DBJ:U84068; NID:q162006; PID:d1012885; PID:q1620007
A:Experimental source: C57 cell
C:Comment: This protein is required for the O-acetylation of disialoganglioside sial1
C:Genetics:
A:Gene: ags
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
F:1-21/Domain: signal sequence #status predicted <Sig>
F:28-60/Domain: EGF homology <EG1>
F:68-107/Domain: EGF homology <EG2>
F:110-267/Domain: discoidin I amino-terminal homology <DN1>
F:271-427/Domain: discoidin I amino-terminal homology <DN2>
Db 947 ISNKMHAINGKMGFNLGGLTHMGDEVVMYVAMGNEDIDLTVFHGHGSHFOYKRGHS 1006
QY 1098 MALYNXPGEVTEVEMLPKAGIMRYECLIGELHAGMSTFLVYSNK 1145
LQ 1007 SDVFDFPGTYOTLEMPPTGTWLLHCHVTDHAGMWTYVLPNQ 1054
RESULT 7
M44258
factor VIII-associated gene B hypothetical protein - human
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 31-Oct-1997
C:Accession: A44258
R:Levenson, B.; Kenwick, S.; Gamel, P.; Fisher, K.; Gitschier, J.
Genomics 14, 585-589, 1992
A:Title: Evidence for a third transcript from the human factor VIII gene.
A:Reference number: A44258; MUID:93052386
A:Accession: A44258
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-216 <LEV>
A:Cross-references: GB:M90707; NID:q182316; PID:q182317
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
F:1-53/Domain: discoidin I amino-terminal homology #status atypical <DN1>
F:57-210/Domain: discoidin I amino-terminal homology <DN2>
Query Match 14.2%; Score 1104; DB 2; Length 216;
Best Local Similarity 100.0%; Pred. No. 9.3e-69; Indels 0; Gaps 0;
Matches 208; Conservative 0; Mismatches 0;
QY 1250 VFGNVDSGIRKININPPIARIYRLPHVYSIRSTLRMELMGCDLNSCMLGESKA 1309
Db 9 VFGNVDSGIRKININPPIARIYRLPHVYSIRSTLRMELMGCDLNSCMLGESKA 68
QY 1310 ISDAQITASSYFTNMFATMSPSKARLHLQGRSNAMRPVNNPKEMLOVDFOKTMKTYGY 1369
Db 69 ISDAQITASSYFTNMFATMSPSKARLHLQGRSNAMRPVNNPKEMLOVDFOKTMKTYGY 128
QY 1370 TQGVKSLTSMYKEFLISSQDGHQWTLTFQNGKAVFQGNDSFTPVNSLDPPLLR 1429
Db 129 TQGVKSLTSMYKEFLISSQDGHQWTLTFQNGKAVFQGNDSFTPVNSLDPPLLR 188
QY 1430 YLRHPSQVWHQIALRMEVYGCDAODLY 1457
Db 189 YLRHPSQVWHQIALRMEVYGCDAODLY 216
RESULT 8
JC4915
ags protein precursor - rat
N:Alternate names: O-acetyl-Gd3 ganglioside
C:Species: Rattus norvegicus (Norway rat)

C:Date: 26-Sep-1996 #sequence_revision 01-Nov-1996 #text_change 13-Nov-1998
C:Accession: JC4915
B:Ogura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanai, Y.
Biochem. Biophys. Res. Commun. 225, 932-938, 1996
A:Title: Cloning and expression of cDNA for O-acetylation of GD3 ganglioside.
A:Reference number: JC4915; MUID:96374422
A:Accession: JC4915
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-427 <OGU>
A:Cross-references: DBJ:U84068; NID:q162006; PID:d1012885; PID:q1620007
A:Experimental source: C57 cell
C:Comment: This protein is required for the O-acetylation of disialoganglioside sial1
C:Genetics:
A:Gene: ags
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
F:1-21/Domain: signal sequence #status predicted <Sig>
F:28-60/Domain: EGF homology <EG1>
F:68-107/Domain: EGF homology <EG2>
F:110-267/Domain: discoidin I amino-terminal homology <DN1>
F:271-427/Domain: discoidin I amino-terminal homology <DN2>
Query Match 8.5%; Score 663; DB 2; Length 427;
Best Local Similarity 40.3%; Pred. No. 5.3e-38;
Matches 142; Conservative 61; Mismatches 123; Indels 26; Gaps 6;
QY 1124 ECLIGELHAGMSTFLVYSNK-----CQPLGASGHIRPQTASGQY-- 1169
Db 78 KCLVTEEDIQRG--DIFTEICCPVGYSGIHCELCSTRLGEGGALDSQISASVYWG 135
QY 1169 ----COMAPKRLRLHYSGSINAW--STKEPFMIKIDLLAPITIGIKQGRKFSSLY 1222
Db 136 FMGLDRMGPELRLKRTGYVNAWMTASIDSCKWIDYDFLRKRVGCVMTQGSRRGRATY 195
QY 1223 ISOFTIMSLSQDKKWQYRNGNSTGLMVFVGVNDSSGIRKININPPIARIYRLPHVYS 1282
Db 196 LKTFKVAYSLDRRREFIIDESGTGDKRFMGQDNNSLKINNFPLEQYIRLVPVSC 255
QY 1283 ISTLRMEIMGCDLNSCMLGESKAISDAQITASSYFT--NMAA-TWSPKARLHLQ 1339
Db 256 RQCTLRFEELGCELGCSSPGLKNNITPDSQITASSSYKTWNLAFGWYPLHGRLDNQG 315
QY 1340 RSNAMPQVNNKREMLQVDFQKTMKTYGVTQGVKSLTSMYKEFLISSQDGHQWTLF 1399
Db 316 KINAWTAQSNASKEMLOVDLGTQKVTGILITQGADFGHIGQVYASRYKVAHSDGQWMTY 375
QY 1400 FQNGKAVFQGNDSFTPVNSLDPPLLTRYLRIHPSQVWHQIALRMEVYGC 1451
Db 376 EEOGTSKYVQGNLNNSHKKNIFERPFMARVYRVLPSWNRHITRLLELGC 427
RESULT 9
A36479
milk fat globule membrane protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change 07-Aug-1998
C:Accession: A36479
R:Stubbbs, J.D.; Lekutis, C.; Singer, K.L.; But, A.; Yuzukl, D.; Srinivasan, U.; Parry
Proc. Natl. Acad. Sci. U.S.A. 87, 8417-8421, 1990
A:Title: cDNA cloning of a mouse mammary epithelial cell surface protein reveals the
A:Reference number: A36479; MUID:91046008
A:Accession: A36479
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-463 <STU>
A:Cross-references: GB:M38337; NID:q199142; PID:q199143
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
F:28-60/Domain: EGF homology <EG1>
F:68-107/Domain: EGF homology <EG2>
F:147-303/Domain: discoidin I amino-terminal homology <DN1>
F:307-463/Domain: discoidin I amino-terminal homology <DN2>

Query Match
Best Local Similarity 8.1%; Score 635; DB 2; Length 427;
Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;

OY 1169 ----GQAPKRLARLHYSSINAMST----KEPFWIKVLDLAPIITIGKITQGAROKRFSS 1220
|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 108 FMGLORMAPELARLHQGTIVNAWTSNGYDKNP--WIDVNLRRKKMWGVGYTGGASRAGSA 165
OY 1221 LYISOFIMYSLDECKMQOTYRGNSGTGLTMFVEGNVDSSGIKHNFNPIIARYLRPLPTH 1280
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 166 EYLKTFKVAYSTDRQRQFIQVAGRSGDKITIGVNNSGKLINFDPPLPEQYRVLPPII 225
OY 1281 YSINTLMELMGCDLNSCSMPLGESKAISDAQITASSVFETN--MFATWSPSKARLHL 1337
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 226 CHRGTLEFFELLGCELNCCTEPLGLKDWTIPNKQITASSSYKKTWGLSASFWEFYARLDN 285
OY 1338 OGRSNAMPQVNNRKREMLQVDFOKTMAKYTYTTQGVSLSLTSMVKEFLISSODGHQMT 1397
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 286 QGKFEMAAWTAQNTNSASEMDLIDLGOSKRVTGTLTGARGFRCHIOYVAAYRYAVGDGVWTI 345
OY 1398 LEFGNG--KVAVFOGNDSPFPVNVSLDPLLTLYRLRHPSWHOTALREVLCG 1451
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 346 EYKDPGASESEKIFPGNMNDNSHKKNIEFTTPQAFVRPIQPAWMHNRILTRVELLGC 401

RESULT 11

PAS-6/7 protein precursor - bovine
S74211
C:Species: Bos primigenius taurus (cattle)
C>Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 01-Feb-1999
A:Accession: S74211; S78114; S24181
R:Hvarregaard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.
Eur.J. Biochem. 240, 628-636, 1996
A>Title: Characterization of glycoprotein PAS-6/7 from membranes of bovine milk fat g
A:Reference number: S74211; MUID:97008954
A:Accession: S74211
A:Molecule type: mRNA
A:Residues: 1-427 <HYVA>
A:Cross-references: EMBL:X91895; NID:g1632778; PID:e204102; PID:g1632779
A:Accession: S78114
A:Molecule type: protein
A:Residues: 19-85;96-110;140-165;174-216;221-232;248-277;285-293;309-337;339-420;425-
R:Kim, D.H.; Kanno, C.; Mizokami, Y.
Biochim. Biophys. Acta 1122, 203-211, 1992
A>Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, f
A:Reference number: S23926; MUID:92353107
A:Accession: S24181
A:Molecule type: protein
A:Residues: 383-394 <KIM>
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
C:Keywords: glycoprotein; milk; blocked amino end; disulfide bond
F:1-18/Domains: signal sequence #status predicted <Sig>
F:19-427/Product: PAS-6/7 protein #status experimental <ANT>
F:6-6/Domain: EGF homology <EGF1>
F:66-105/Domains: EGF homology <EGF2>
F:108-265/Domains: discoidin I amino-terminal homology <DN1>
F:265-427/Domains: discoidin I amino-terminal homology <DN2>
F:24-35;29-47;49-58;66-77;71-94;96-105/Disulfide bonds: #status predicted
F:27/Binding site: carbohydrate (Ser) (covalent) #status experimental
F:34/Binding site: carbohydrate (Thr) (covalent) #status experimental
F:59;122/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:109-265;252-256;270-427/Disulfide bonds: #status experimental

```

      192 EYLKFKVAVITDGRQRFQIVAGSGSKITIGVNNNGCLINLEDFPLEYQYRLVPTI 251
      QY 1281 YSINSTRMELMGCDLNSCSMPLGNESKAISDAQITASSYTTN--MFATWSPSKARLHL 1337
      Db 252 CHRGTLLFELLGCLNCTEPLGLKDMTIPNKQITASSYYKTWGLSAFSEFPYARLDN 311
      QY 1338 OGRSNAMPVQNNPREKMLQVDFQKTMKVTGTGKSLTSMVYKEFLISSODGHW 1397
      Db 312 OGRFAMTAQITASSEMQLDIDGSOGRVYGLITGARGHGIQYAAVRAVAGDDGVMT 371
      QY 1398 LEFONG--KVFVQGNODSFTEPVVNSLDPLRLTRLRHPSWQIALRMEVLGC 1451
      Db 372 EYKDGASBESKIFPPNMNNSHKNIFFEPFGARFVRILQPVAMHNRILRLVELLGC 427

```

RESULT 12

```

      945
      C:Date: 16-Aug-1998 #sequence_revision 16-Aug-1998 #text_change 31-Dec-1996
      C:Accession: A25945
      R:TOOLE, J.J.; Pittman, D.D.; Orr, E.C.; Murtha, P.; Wasley, L.C.; Kaufman, R.J.
      Proc. Natl. Acad. Sci. U.S.A. 83, 5939-5942, 1986
      A:Title: A large region (approx 95 kDa) of human factor VIII is dispensable for in vitro
      A:Reference number: A25945; MUID:86287369
      A:Accession: A25945
      A:Status: nucleic acid sequence not shown; not compared with conceptual translation
      A:Molecule type: DNA
      A:Residues: 1-869 <T00>

```

```

      Query Match
      Best Local Similarity 75.2%; Pred. No. 2.9e-29;
      Matches 100; Conservative 14; Mismatches 19; Indels 0; Gaps 0;
      QY 765 PVLKRRHREITRTLLQSDQEIYDDTISVEKKEDFDIYDEDNQSRSPQKTRHYF 824
      Db 737 PVLKRRHREITRTLLQSDQEIYDDTISVEKKEDFDIYDEDNQSRSPQKTRHYF 796
      QY 825 IAAVEKLMYGMSSPHVLRNRAQSGSVQFQKVVYQFQETDQSFQPLRGELNEHLGL 884
      Db 797 IAAVEKLMYGMSSPHVLRNRAQSGSVQFQKVVYQFQETDQSFQPLRGELNEHLGL 856
      QY 885 GPYIRAEVEDNIM 897
      Db 857 GPYIRAEVEDNIM 869

```

RESULT 13

```

      J00948
      A:Antigen precursor - African clawed frog
      C:Species: Xenopus laevis (African clawed frog)
      C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Mar-1998
      C:Accession: J00948
      R:Itakagi, S.; Hirata, T.; Agata, K.; Mochii, M.; Eguchi, G.; Fujisawa, H.
      Neuron 7, 295-307, 1991
      A:Title: The A5 antigen, a candidate for the neuronal recognition molecule, has homology
      A:Reference number: J00466; MUID:91337458
      A:Accession: J00466
      A:Molecule type: mRNA
      A:Residues: 1-927 <TAK>
      A:Cross-references: GB:D10467; GB:D01077; NID:g222966; PID:d1001730; PID:g222963
      A:Experimental source: tadpole, brain
      A:Note: This protein has motifs homologous to complement components C1r and C1s and to C
      C:Superfamily: Xenopus A5 antigen; C1r/C1s repeat homology; discoidin I amino-terminal h
      C:Keywords: duplication; glycoprotein; transmembrane protein
      F:1-21/Domain: signal sequence #status predicted <Sig>
      F:22-927/Product: A5 antigen #status predicted <A5A>
      F:27-138/Domain: C1r/C1s repeat homology <C1R1>
      F:147-262/Domain: C1r/C1s repeat homology <C1R2>
      F:274-424/Domain: discoidin I amino-terminal homology <DN1>

```

```

      F:430-584/Domain: discoidin I amino-terminal homology <DN2>
      F:646-812/Domain: NAM homology <NAM>
      F:861-883/Domain: transmembrane #status predicted <TMA>
      F:150,261,300,523,844/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

```

      Query Match
      Best Local Similarity 29.9%; Pred. No. 2.3e-22;
      Matches 138; Conservative 75; Mismatches 168; Indels 80; Gaps 19;

```

```

      QY 1065 IRTVLSMGSNENHSHFSGHVFYAKKEEYKMAIYNLY-----PGV-----FEVEM- 1114
      Db 136 IRYEVEFTGECSKNTSSNGVKSIPKYPKIPNALECTYIIPAKQOEIVLEFESELE 195
      QY 1114 LPKRA-----GIMVECLIGEHL--HAGMST----- 1138
      Db 196 ADSNAPGCGCRMYDGLIMGDFGVGHIRYGGQNPGRVRSPTGILSMIFHDSAIK 255
      QY 1138 -----LELYSN-----KQPLGMAASHIRDFQITASQYQ-ONAPKILRLHSGSIN 1185
      Db 256 EGFPANFSVQSNTEDEDFQCKEALGMEGSHFDQLSVSSQYSNMMSAERSRLNYE--N 313
      QY 1186 AMSKKEP--SMIKVDLAPMIHIGIKTGA--RQKSSLYISOEIIIMYSIDGKKQYR 1241
      Db 314 GMTGDETKEMIDYDLNLFVSGITQAIKSETKKTFVSKYKVDISNGEDWITLK 373
      QY 1242 -GNSTGLMVFEGNVDSGKIKNIFNPPIIARYIRLPHYSIRSTRLMELMGCDLNS-- 1299
      Db 374 DGNH---LVFTGMTADTVYRPFSPVIRFRLRPVWENGISIRFELYGCKITDYP 430
      QY 1299 CSMPLGMEKRAISDAQITASSYFTNMFATWSPSKARLHLGSRNAMPVQNNP--KEMLO 1356
      Db 431 CSRMIGVNSGLISDSQITASS--QVDRNWPBELARLVTSRSGMALPSPNHPPTKEMLO 487
      QY 1357 VDFQTKMVTGVTGQVKSLLTSMYKEFLISSQSDGHWTLFQN--GKVKYQGNQDS 1414
      Db 488 IDLAEKIVRGLIIOGKHKNKVFRKFGYSNNGEMIMDSKPKKFEGETNTY 547
      QY 1415 FTPVANSIDPLRLTRLRHPS--SWHQIALRMEVLGCEAO 1454
      Db 548 DPELRLTF-AHITGFIIRIIPERASAGIALRLLELCEVE 587

```

RESULT 14

```

      A47285
      C:Species: Homo sapiens (man)
      C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 31-Oct-1997
      C:Accession: A47285
      R:Latocca, D.; Peterson, J.A.; Ureia, R.; Kunjoshi, J.; Bistrain, A.M.; Ceriani, R.L.
      Cancer Res. 51, 4994-4996, 1991
      A:Title: A M-r 46,000 human milk fat globule protein that is highly expressed in huma
      A:Reference number: A47285
      A:Accession: A47285
      A:Status: preliminary
      A:Molecule type: mRNA
      A:Residues: 1-218 <LAP>
      C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
      F:1-56/Domain: discoidin I amino-terminal homology (fragment) <DN1>
      F:60-218/Domain: discoidin I amino-terminal homology <DN2>

```

```

      Query Match
      Best Local Similarity 40.6%; Pred. No. 4.5e-22;
      Matches 84; Conservative 44; Mismatches 72; Indels 7; Gaps 3;

```

```

      QY 1251 FPGVDSGKIKINIPPIIARYIRLPHYSIRSTRLMELMGCDLNSCSMPLGESKAI 1310
      Db 13 FPGNNKNAVHNLEETPEAQYVRLPYTSCTACTLRFELGCLNGCANPLGKNNSI 72
      QY 1311 SDAQITASSYF---TNMFATWSPSKARLHLGSRNAMPVQNNKEMLOVDFQTKMVT 1366
      Db 73 PKQITASSSYKTWGLHIF-SWNPAYARLDKQGNFNAVAGSYGNDWLOLDLGSSKEVT 131

```


GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 1999, 12:00:26 ; Search time 12.92 Seconds

(without alignments)
3187,841 Million cell updates/sec

Title: **US-09-001-039a-47**
Perfect score: 7797
Sequence: 1 M0E1STCFPLLCRLRCFSA.....WVHQAIRMEVLCECAEDLY 1457

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database: **SwissProt_37**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7340	94.1	2351	1	FA8_HUMAN
2	6306	80.9	2133	1	FA8_PIG
3	6261	80.3	2139	1	FA8_MOUSE
4	2390.5	30.7	2224	1	FA5_HUMAN
5	2370	30.4	2211	1	FA5_BOVIN
6	1741	22.3	1065	1	CERU_HUMAN
7	1694	21.7	1059	1	CERU_RAT
8	1651	21.2	1062	1	CERU_MOUSE
9	663	8.3	427	1	MEGM_MOUSE
10	657	8.4	463	1	MEGM_MOUSE
11	650	8.3	409	1	MEGM_PIG
12	635	8.1	427	1	MEGM_BOVIN
13	588	7.5	387	1	MEGM_HUMAN
14	458.5	5.9	914	1	NRP_CHICK
15	446.5	5.7	923	1	NRP_MOUSE
16	443	5.9	928	1	NRP_XENLA
17	306.5	3.9	3133	1	HMC1_BOVMO
18	198.5	2.5	854	1	TRK3_MOUSE
19	192	2.5	913	1	EDD1_HUMAN
20	185.5	2.4	911	1	EDD1_MOUSE
21	185.5	2.4	910	1	EDD1_RAT
22	185.5	2.4	855	1	TRK3_HUMAN
23	149.5	1.9	578	1	ASO_TOBAC
24	148	1.9	622	1	YAK8_SCHPO
25	140.5	1.8	1196	1	BXCN_CLOBO
26	130	1.7	818	1	DAP2_YEAST
27	130	1.7	4196	1	DYHC_SCHPO
28	130	1.7	949	1	HISA_STRPN
29	128.5	1.6	994	1	DPO2_KLJUA
30	125	1.6	2167	1	BEM2_YEAST
31	125	1.6	2368	1	ESR1_YEAST
32	123	1.6	387	1	ASO_CUCSA
33	123	1.6	1471	1	MISA_YEAST
34	123	1.6	867	1	SSPO_BOVIN
35	121.5	1.6	2216	1	YCF2_EPIVI
36	121.5	1.6	2280	1	YCF2_TORAC
37	120.5	1.5	1405	1	DPOA_SCHPO
38	119.5	1.5	1098	1	RPOP_MAIZE
39	119.5	1.5	1648	1	YJ9H_YEAST
40	119	1.5	1165	1	LEPR_HUMAN
41	118.5	1.5	923	1	PWP2_YEAST
42	118.5	1.5	2167	1	YCS2_YEAST
43	117.5	1.5	1430	1	GRFD_STRMU

ALIGNMENTS

RESULT	1	44	117.5	1.5	1139	1	HMN1_MYGE
FA8_HUMAN		45	117.5	1.5	1461	1	TOP2_CANAL
FA8_HUMAN	STANDARD;						
P00451;	PRT;						
AC	21-JUL-1986 (REL. 01, CREATED)						
DT	21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)						
DT	15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)						
DE	COAGULATION FACTOR VIII PRECURSOR (PROCOAGULANT COMPONENT).						
GN	F8C.						
OS	HOMO SAPIENS (HUMAN).						
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;						
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE; 86081164.						
RA	TRETT M.A., BLACHER R., BURKE R.L., CAPUT D., CHU C., DINA D.,						
RA	HARTOG K., KUO C.H., MASTARZ F.R., MERRYWEATHER J.P., NAJARIAN R.,						
RA	PACHL C., POTTER S.J., PODA J., QUITROGA M., RALL L.B., RANDOLPH A.,						
RA	URDEA M.S., VALENZUELA P., DAHL H.-H.M., FAVALLARO J., HANSEN J.,						
RT	"Characterization of the polypeptide composition of human factor						
RT	viii:c and the nucleotide sequence and expression of the human kidney						
RT	cDNA."						
RL	DNA 4:333-349(1985).						
RN	[2]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE; 85061548.						
RA	WOOD W.I., CAPON D.J., SIMONSEN C.C., EATON D.L., GITSCHIER J.,						
RA	KEYT B., SEEBURG P.H., SMITH D.H., HOLLINGSHEAD P., WION K.L.,						
RA	DELMAT E., TUDENHAM E.G.D., VEHR G.A., LAMN R.M.,						
RT	"Expression of active human factor VIII from recombinant DNA clones."						
RT	NATURE 312:330-337(1984).						
RN	[3]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE; 85061550.						
RA	TOOLE J.T., KNOFF J.L., WOZNEY J.M., SULTZMAN L.A., BUECKER J.L.,						
RA	PITTMAN D.D., KAUFMAN R.J., BROWN E., SHOEMAKER C., ORR E.C.,						
RA	AMPHLETT G.W., FOSTER W.B., COE M.L., KNUITSON G.J., FASS D.N.,						
RA	HEWICK R.M.,						
RT	"Molecular cloning of a cDNA encoding human antithrombin factor."						
RT	NATURE 312:342-347(1984).						
RL	[4]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE; 93265012.						
RA	GITSCHIER J., WOOD W.I.,						
RT	"Sequence of the exon-containing regions of the human factor VIII						
RT	gene."						
RL	HUM. MOL. GENET. 1:199-200(1992).						
RN	[5]						
RP	SEQUENCE OF 2064-2070 FROM N.A.						
RA	DE WATER N.S., WILLIAMS R., BROWETT P.J.,						
RL	SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.						
RN	[6]						
RP	SULFATATION OF 1699.						
RX	MEDLINE; 91093266.						
RA	LEYTE A., VAN SCHIJNDEL H.B., NIEHRS C., HUTTNER W.B., VERBEET M.P.,						
RA	MERTENS K., VAN MOORIK J.A.,						
RT	"Sulfation of Tyr1680 of human blood coagulation factor VIII is						
RT	essential for the interaction of factor VIII with von Willebrand						
RT	factor."						
RL	J. BIOL. CHEM. 266:740-746(1991).						
RN	[7]						
RP	SULFATATION.						
RX	MEDLINE; 92207952.						
RA	PITTMAN D.D., WANG J.H., KAUFMAN R.J.,						
RT	"Identification and functional importance of tyrosine sulfate						

RT residues within recombinant factor VIII.";
 RL BIOCHEMISTRY 31:3315-3325(1992).
 RN [8]
 RP STRUCTURE BY NMR OF 2322-2343.
 RX MEDLINE: 95200924.
 RA GILBERT G.E., BALEJA J.D.;
 RT "Membrane-binding peptide from the C2 domain of factor VIII forms an
 RT amphipathic structure as determined by NMR spectroscopy.";
 RL BIOCHEMISTRY 34:3022-3031(1995).
 RN [9]
 RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
 RX MEDLINE: 91221499.
 RA GITSCHER J.;
 RT "The molecular basis of hemophilia A.";
 RL ANN. N.Y. ACAD. SCI. 614:89-96(1991).
 RN [10]
 RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
 RX MEDLINE: 89088506.
 RA WHITE G.C. II, SHOEMAKER C.B.;
 RT "Factor VIII gene and hemophilia A.";
 RL BLOOD 73:1-12(1989).
 RN [11]
 RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
 RX MEDLINE: 95245332.
 RA ANTONARAKIS S.E., KAZAZIAN H.H., TUDDENHAM E.G.D.;
 RT "Molecular etiology of factor VIII deficiency in hemophilia A.";
 RL HUM. MUTAT. 5:1-22(1995).
 RN [12]
 RP VARIANT GLN-2326.
 RX MEDLINE: 86235434.
 RA GITSCHER J., WOOD W.I., SHUMAN M.A., LAMN R.M.;
 RT "Identification of a missense mutation in the factor VIII gene of a
 RT mild hemophiliac.";
 RL SCIENCE 232:1415-1416(1986).
 RN [13]
 RP VARIANT PRO-2135.
 RX MEDLINE: 86096539.
 RA LEVINSOHN B., JANCO R.L., PHILLIPS J.A. III, GITSCHER J.;
 RT "A novel missense mutation in the factor VIII gene identified by
 RT analysis of amplified hemophilia DNA sequences.";
 RL NUCLEIC ACIDS RES. 15:9797-9805(1987).
 RN [14]
 RP VARIANT GLN-2228.
 RX MEDLINE: 86191889.
 RA YOUSSEFJIAN H., ANTONARAKIS S.E., BELL W., GRIFFIN A.M.,
 RL KAZAZIAN H.H.;
 RT "Nonsense and missense mutations in hemophilia A: estimate of the
 RT relative mutation rate at CG dinucleotides.";
 RL AM. J. HUM. GENET. 42:718-725(1988).
 RN [15]
 RP VARIANT GLY-291.
 RX MEDLINE: 88220354.
 RA YOUSSEFJIAN H., WONG C., ARONIS S., PLATOKOUKIS H., KAZAZIAN H.H. JR.,
 RT ANTONARAKIS S.E.;
 RT "Moderately severe hemophilia A resulting from Glu-->Gly substitution
 RT in exon 7 of the factor VIII gene.";
 RL AM. J. HUM. GENET. 42:867-871(1988).
 RN [16]
 RP VARIANT CYS-1708.
 RX MEDLINE: 89274393.
 RA O'BRIEN D.P., TUDDENHAM E.G.;
 RT "Purification and characterization of factor VIII 1,689-Cys: a
 RT nonfunctional cofactor occurring in a patient with severe hemophilia
 RT A.";
 RL BLOOD 73:2117-2122(1989).
 RN [17]
 RP VARIANT CYS-391.
 RX MEDLINE: 90001543.
 RA SHIMA M., WARE J., YOSHIOKA A., FUKUI H., FULCHER C.A.;
 RT "An arginine to cysteine amino acid substitution at a critical
 RT thrombin cleavage site in a dysfunctional factor VIII molecule.";
 RL BLOOD 74:1612-1617(1989).
 RN [18]
 RP VARIANT LEU-189.
 RX MEDLINE: 90057680.
 RA CHAN V., CHAN T.K., TONG T.M., TODD D.;
 RT "A novel missense mutation in exon 4 of the factor VIII:C gene
 RT resulting in moderately severe hemophilia A.";
 RL BLOOD 74:2688-2691(1989).
 RN [19]
 RP VARIANT LEU-2326.
 RX MEDLINE: 89197216.
 RA INABA H., FUJIMAKI M., KAZAZIAN H.H. JR., ANTONARAKIS S.E.;
 RT "Mild hemophilia A resulting from Arg-to-Leu substitution in exon 26
 RT of the factor VIII gene.";
 RL HUM. GENET. 81:335-338(1989).
 RN [20]
 RP VARIANT HIS-391.
 RX MEDLINE: 89264602.
 RA ARAI M., INABA H., HIGUCHI M., ANTONARAKIS S.E., KAZAZIAN H.H. JR.,
 RT FUJIMAKI M., HOYER L.W.;
 RT "Direct characterization of factor VIII in plasma: detection of a
 RT mutation altering a thrombin cleavage site
 RT (arginine-372-->histidine).";
 RL PROC. NATL. ACAD. SCI. U.S.A. 86:4277-4281(1989).
 RN [21]
 RP VARIANT CYS-1708.
 RX MEDLINE: 90105723.
 RA ARAI M., HIGUCHI M., ANTONARAKIS S.E., KAZAZIAN H.H. JR.,
 RT PHILLIPS J.A. III, JANCO R.L., HOYER L.W.;
 RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to
 RT Cys) in the factor VIII gene of two unrelated patients with
 RT cross-reacting material-positive hemophilia A.";
 RL BLOOD 75:384-389(1990).
 RN [22]
 RP VARIANTS GLN-2228 AND LEU-2326.
 RX MEDLINE: 90123183.
 RA CASUJA L., MORRU S., PECORARA M., RISTALDI M.S., RESRAGNO G.,
 RA MANCUSO G., MORFINI M., DE BIASI R., BAUDO F., CARBONARA A.;
 RT "Recurrent mutations and three novel rearrangements in the factor
 RT VIII gene of hemophilia A patients of Italian descent.";
 RL BLOOD 75:662-670(1990).
 RN [23]
 RP VARIANT CYS-391.
 RX MEDLINE: 90329422.
 RA PATTINSON J.K., MCVEY J.H., BOON M., AJANI A., TUDDENHAM E.G.;
 RT "CRM+ haemophilia A due to a missense mutation (372-->Cys) at the
 RT internal heavy chain thrombin cleavage site.";
 RL BR. J. HAEMATOL. 75:73-77(1990).
 RN [24]
 RP VARIANTS PHE-1699 AND CYS-1708.
 RX MEDLINE: 90152691.
 RA HIGUCHI M., WONG C., KOCHAN L., OLEK K., ARONIS S., KASPER C.K.,
 RA KAZAZIAN H.H., ANTONARAKIS S.E.;
 RT "Characterization of mutations in the factor VIII gene by direct
 RT sequencing of amplified genomic DNA.";
 RL GENOMICS 6:65-71(1990).
 RN [25]
 RP VARIANTS CYS-1728 AND ASP-1941.
 RX MEDLINE: 90169988.
 RA TRAYSTMAN M.D., HIGUCHI M., KASPER C.K., ANTONARAKIS S.E.,
 RA KAZAZIAN H.H.;
 RT "Use of denaturing gradient gel electrophoresis to detect point
 RT mutations in the factor VIII gene.";
 RL GENOMICS 6:293-301(1990).
 RN [26]
 Query Match 94.1%; Score 7340; DB 1; Length 2351;
 Best Local Similarity 62.0%; Pred. No. 0;
 Matches 1457; Conservative 0; Mismatches 0; Indels 894; Gaps 1;
 OY 1 MOIELSTCFCLLRPCFSATRRYYLGAVELSDWQSDGLGELPYDARPPRPSPFPFN 60
 DB 1 MOIELSTCFCLLRPCFSATRRYYLGAVELSDWQSDGLGELPYDARPPRPSPFPFN 60
 OY 61 TSVYTKKTLFVEFTDHLNIAKPRPPWGLGPTTIOAEYDITVITLKNMASHPVSLNAV 120

Db	61	TSVYKKTLEFEETDHLFENAKRPPMVGGLGTQIAEYDYTVITLKNMASHPSYLAHV	120
QY	121	GVSVFMKASEGALEYDDQTSQREKEDDKFYPCGSGSTHYWQVLKENGPMASDPLCLTFYSYLH	180
Db	121	GVSVFMKASEGALEYDDQTSQREKEDDKFYPCGSGSTHYWQVLKENGPMASDPLCLTFYSYLH	180
QY	181	VDLYKDLNSGLIGALLVCREGSLAKEKTOYTLAKFILLFAVDEGKSWHSETKNSLMODRD	240
Db	181	VDLYKDLNSGLIGALLVCREGSLAKEKTOYTLAKFILLFAVDEGKSWHSETKNSLMODRD	240
QY	241	AASRAHPKMHYTVGVYNRSLPGLJGCHRSYVWHYIGMGTPEVHSTFLECHTFLYVNH	300
Db	241	AASRAHPKMHYTVGVYNRSLPGLJGCHRSYVWHYIGMGTPEVHSTFLECHTFLYVNH	300
QY	301	QOASLEISPTFLTAQTLMLMDLGOFLTFCHISSHQDGEAVYKVDSCPEEQLBPKMKN	360
Db	301	QOASLEISPTFLTAQTLMLMDLGOFLTFCHISSHQDGEAVYKVDSCPEEQLBPKMKN	360
QY	361	EAEYEDDDLDESEMDVYRFDDNSPSFTQIYRSYAKKHPTWYHITAAEEDMDVAPLYLA	420
Db	361	EAEYEDDDLDESEMDVYRFDDNSPSFTQIYRSYAKKHPTWYHITAAEEDMDVAPLYLA	420
QY	421	PDDBSYSOYVYLNNGPORGKRYKKVPRMAYTDETFKTRALIOHESGILGPLYGSEVDTL	480
Db	421	PDDBSYSOYVYLNNGPORGKRYKKVPRMAYTDETFKTRALIOHESGILGPLYGSEVDTL	480
QY	481	LILFKNOASRPYNIYPHGINDVAPLYRSRRLPKGVKHLKDFLILPGEIFKXKTYVEDGP	540
Db	481	LILFKNOASRPYNIYPHGINDVAPLYRSRRLPKGVKHLKDFLILPGEIFKXKTYVEDGP	540
QY	541	TKSDPRCLTRYSSFFVMEKEDLASGLIGPLITCYKSYSOQRGNQIMSDKRVNITLSEVDE	600
Db	541	TKSDPRCLTRYSSFFVMEKEDLASGLIGPLITCYKSYSOQRGNQIMSDKRVNITLSEVDE	600
QY	601	NRSWYLTENIORFLPNBAGVQLEDEPEOASINMHSINGVYFDSLOLQSVCLHEVAAWYILS	660
Db	601	NRSWYLTENIORFLPNBAGVQLEDEPEOASINMHSINGVYFDSLOLQSVCLHEVAAWYILS	660
QY	661	IGAOTDPLSVFSGSYTEFKHKMYEDLTLTFPESGETVFMSEMPGAILGCHNSDFRRNG	720
Db	661	IGAOTDPLSVFSGSYTEFKHKMYEDLTLTFPESGETVFMSEMPGAILGCHNSDFRRNG	720
QY	721	MTALIKYSCCKNTGDXEYEDYSIEDISAYLLSKNNALPESP-----	762
Db	721	MTALIKYSCCKNTGDXEYEDYSIEDISAYLLSKNNALPESP-----	762
QY	762	-----	762
Db	762	-----	762
QY	841	PGAIDSNNSISEMTHFRPOLHSGDMVTFPESGLQRLNEKLGTTPATATLKLDFKVSST	900
Db	841	PGAIDSNNSISEMTHFRPOLHSGDMVTFPESGLQRLNEKLGTTPATATLKLDFKVSST	900
QY	901	SNNLISTIPSDNLAAGTDNNTSSLGPPSPMPVHYDSOLDLTLFGKKSPLTESGGLPILSEE	960
Db	901	SNNLISTIPSDNLAAGTDNNTSSLGPPSPMPVHYDSOLDLTLFGKKSPLTESGGLPILSEE	960
QY	961	NNDSKLTLESGLMNOESBGMKKNVSTSESGLFPGRKAHGPALLTKNALFKVYSILATN	1020
Db	961	NNDSKLTLESGLMNOESBGMKKNVSTSESGLFPGRKAHGPALLTKNALFKVYSILATN	1020
QY	762	-----	762
Db	1021	KTSNNSATNRKTHIDGPSLIENSPSYWONILESDTEFKKVTPLIHDRMLMDKNATLRL	1080
QY	762	-----	762
Db	1081	NHMSKTKTSSKNMNMVOQKKEGPILPPAQNPDMSFKMLFLPESANWIORTHGKNSLNSG	1140
QY	762	-----	762

Db	1141	QGSPPKQVLISLPEKSVGEQNFISEKKNVVGSGEFTKDVGLKEKVEPSSRNLEFTLNLD	1200
Qy	762	-----	762
Db	1201	LHENNTINQEKKIOEIEKEKETLIOENVVLPOIHVYTGKKNMKMLFLSTRQWVEGSD	1260
Qy	762	-----	762
Db	1261	GAYAPVLQDFRSLNDSTNPTFKHTAHFSKKEBENLEGLNQTKQIVEKYACTTRISPNT	1320
Qy	762	-----	762
Db	1321	SOQNVYTORSKRALQFRLPLEEFTELEKRIIVDDTSTQSKNMKHLPTSTLIOIDYNEKE	1380
Qy	762	-----	762
Db	1381	KGALTQSPLSCLTRSHSTIQANRSPPLAKVSSPFSIRPYLTLYVLEQDNSSHLPAA5Y	1440
Qy	762	-----	762
Db	1441	RKKGQVQESHFLQGAKKNNLSLALITLMTGDQREVQSLGTSATNSVTYKKEVNTLP	1500
Qy	762	-----	762
Db	1501	KPDLPTSGKVELLPRVHIYQXDLPEFTETSNQSPGHLDLVEGSLQGTGEGAIKWNANRP	1560
Qy	762	-----	762
Db	1561	GKVPPLKVAIESSAKTPTSKLLDPLANDNHQIOIPREEMKSOEKSPEXTARKKDTIISL	1620
Qy	762	-----SONPVLKRHQREITRTTLOSQDE -----	762
Db	1621	NACESNHAIAINEGQKKPELEVYMAKQGTHERLCSQNPVLRKHQREITRTTLOSQDE	1680
Qy	767	IDYDDTISVEKKKEDPITYEDENQSPRSFOKTRIFYAAVERLMDYGMSSPHVLRNR	846
Db	1681	IDYDDTISVEKKKEDPIYEDENQSPRSFOKTRIFYAAVERLMDYGMSSPHVLRNR	1740
Qy	847	AQSGSVPOFKVVOEFTDGSFTOPYRGELNHLGLPTRYAEVENIMVTFPNQSR	906
Db	1741	AQSGSVPOFKVVOEFTDGSFTOPYRGELNHLGLPTRYAEVENIMVTFPNQSR	1800
Qy	907	PYSFYSLSI5YEDDQROGAPRRKNFYKPNYETKIYFKVQHHNAPTKDEDFCKAMAYFSDV	966
Db	1801	PYSFYSLSI5YEDDQROGAPRRKNFYKPNYETKIYFKVQHHNAPTKDEDFCKAMAYFSDV	1860
Qy	967	DLEKDVHSLGILPLVCHTNTLPANRGROYVOEFALEFTTIDETKSWYFENNERNCRA	1026
Db	1861	DLEKDVHSLGILPLVCHTNTLPANRGROYVOEFALEFTTIDETKSWYFENNERNCRA	1920
Qy	1027	PCNTQMEPPTFKENYRPHALNGYIMDTLGLVAAOQRIHWLLSMGSENENHSHFEGH	1086
Db	1921	PCNTQMEPPTFKENYRPHALNGYIMDTLGLVAAOQRIHWLLSMGSENENHSHFEGH	1980
Qy	1087	VETVAKKEEYKFMALYNXPQVFETVEMLPSKAGIMRVECLIGEHJHAGSTLEFYNSKC	1146
Db	1981	VETVAKKEEYKFMALYNXPQVFETVEMLPSKAGIMRVECLIGEHJHAGSTLEFYNSKC	2040
Qy	1147	QTPJGMSAGHIRDFQITASQOYGOMAPKLARLYSGSINAMSTKEPFSWIKVDLAPMI	1206
Db	2041	QTPJGMSAGHIRDFQITASQOYGOMAPKLARLYSGSINAMSTKEPFSWIKVDLAPMI	2100
Qy	1207	HGKIQGARQKQFSSLYISQTTIMSLDGKKWQTYRGNSSTGTLMAVFFGVNDSSGIKHNITN	1266
Db	2101	HGKIQGARQKQFSSLYISQTTIMSLDGKKWQTYRGNSSTGTLMAVFFGVNDSSGIKHNITN	2160
Qy	1267	PIIARVYRLRPHYYSIRSLRMEJLWGCDLNSC5MPLGME5KAISDAQITASSYPTN5EA	1326
Db	2161	PIIARVYRLRPHYYSIRSLRMEJLWGCDLNSC5MPLGME5KAISDAQITASSYPTN5EA	2220
Qy	1327	TWSPSKARLHLQGSNAMRPOVNNPKEMLOVDFQKTMKYVTGYTQGVK5LLTSMYKBEFL	1386
Db	2221	TWSPSKARLHLQGSNAMRPOVNNPKEMLOVDFQKTMKYVTGYTQGVK5LLTSMYKBEFL	2280

Oy	1387	ISSSDGHQWTFEPONGKVKYQGNDSFTPVNSLDPILTRRYLRHPQSHVQHTALRM	1446
Db	2281	ISSSDGHQWTFEPONGKVKYQGNDSFTPVNSLDPILTRRYLRHPQSHVQHTALRM	2340
Qy	1447	EVLGCEADPLY 1457	
Db	2341	EVLGCEADPLY 2351	
RESULT	2		
FAB_PIG	STANDARD:	PRT: 2133 AA.	
AC	P12263:	Q95243:	
DT	01-OCT-1989	(REL. 12, CREATED)	
DT	01-NOV-1987	(REL. 35, LAST SEQUENCE UPDATE)	
DT	01-NOV-1997	(REL. 35, LAST ANNOTATION UPDATE)	
DT	COAGULATION FACTOR VIII PRECURSOR (PROCOAGULANT COMPONENT), CFB.		
OC	SUS SCROFA (PIG).		
OC	EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: MAMMALIA: EUTHERIA:		
OC	ARTIODACTYLA: SUIFORMES: SUINA: SUIDAE: SUS.		
RA	[1]		
RA	SEQUENCE FROM N.A.		
RA	HEALEY J.F., LUBIN I.M., LOLLAR P.;		
RL	SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.		
RP	[2]		
RP	SEQUENCE OF 705-1573 FROM N.A.		
RX	MEDLINE; 86287369.		
RA	TOOLE J.J., PITTMAN D.D., ORR E.C., MURTHA P., WASLEY L.C.,		
RA	KAUFMAN R.J.;		
RT	"A large region (approximately equal to 95 kDa) of human factor VIII		
RT	is dispensable for in vitro procoagulant activity.";		
RL	PROC. NATL. ACAD. SCI. U.S.A. 83:5939-5942(1986).		
RN	[3]		
RP	SEQUENCE OF 392-759 FROM N.A.		
RX	MEDLINE; 94179260.		
RA	LUBIN I.M., HEALEY J.F., SCANDELLA D., RUNGE M.S., LOLLAR P.;		
RT	"Elimination of a major inhibitor epitope in factor VIII.";		
RL	J. BIOL. CHEM. 269:8639-8641(1994).		
CC	-I- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS		
CC	AS A COFACTOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE		
CC	ACTIVATED FORM, FACTOR XA.		
CC	-I- SUBCELLULAR LOCATION: EXTRACELLULAR.		
CC	-I- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF		
CC	2 PLASTOCYANIN-LIKE REPEATS.		
CC	-I- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.		
CC	-I- SIMILARITY: STRONG, TO COAGULATION FACTOR V.		
CC	-----		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on list-		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL: U49517; G1511634; -.		
DR	PIR: A25945; A25945.		
DR	PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.		
DR	PROSITE; PS01285; FA58C_1; 2.		
DR	PROSITE; PS01286; FA58C_2; 2.		
DR	PFAM; PF00394; Cu-oxidase; 3.		
DR	PFAM; PF00754; F5_F8_type_C; 2.		
DR	HSP; P00451; ICRG.		
DR	BLOOD COAGULATION; REPEAT; PLASMA; ACUTE PHASE; CALCIUM;		
RW	SIGNAL. GLYCOPROTEIN; SULFATATION.		
FT	SIGNAL. 1 19	POTENTIAL.	
FT	CHAIN 20 2133	COAGULATION FACTOR VIII.	
FT	DOMAIN 20 357	F5/8 TYPE A 1.	
FT	DOMAIN 20 199	PLASTOCYANIN-LIKE 1.	
FT	DOMAIN 207 357	PLASTOCYANIN-LIKE 2.	
FT	DOMAIN 399 730	F5/8 TYPE A 2.	

FT	DOMAIN	399	573		PLASTOCYANIN-LIKE 3.
FT	DOMAIN	583	730		PLASTOCYANIN-LIKE 4.
FT	DOMAIN	760	1599		B.
FT	DOMAIN	1495	1822		F5/8 TYPE A 3.
FT	DOMAIN	1495	1659		PLASTOCYANIN-LIKE 5.
FT	DOMAIN	1669	1822		PLASTOCYANIN-LIKE 6.
FT	DOMAIN	1822	1970		F5/8 TYPE C 1.
FT	DOMAIN	1975	2127		F5/8 TYPE C 2.
FT	SITE	391	392		CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT	SITE	759	760		CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT	SITE	1449	1450		CLEAVAGE (ACTIVATION) (BY SIMILARITY).
FT	SITE	1490	1491		CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT	MOD_RES	737	737		SULFATATION (BY SIMILARITY).
FT	MOD_RES	738	738		SULFATATION (BY SIMILARITY).
FT	MOD_RES	742	742		SULFATATION (BY SIMILARITY).
FT	DISULFID	173	199		PROBABLE.
FT	DISULFID	547	573		PROBABLE.
FT	DISULFID	1633	1659		PROBABLE.
FT	DISULFID	1822	1970		BY SIMILARITY.
FT	DISULFID	1975	2127		BY SIMILARITY.
FT	CARBOHYD	233	233		POTENTIAL.
FT	CARBOHYD	259	259		POTENTIAL.
FT	CARBOHYD	601	601		POTENTIAL.
FT	CARBOHYD	929	929		POTENTIAL.
FT	CARBOHYD	985	985		POTENTIAL.
FT	CARBOHYD	1025	1025		POTENTIAL.
FT	CARBOHYD	1111	1111		POTENTIAL.
FT	CARBOHYD	1181	1181		POTENTIAL.
FT	CARBOHYD	1208	1208		POTENTIAL.
FT	CARBOHYD	1245	1245		POTENTIAL.
FT	CARBOHYD	1265	1265		POTENTIAL.
FT	CARBOHYD	1335	1335		POTENTIAL.
FT	CARBOHYD	1408	1408		POTENTIAL.
FT	CARBOHYD	1611	1611		POTENTIAL.
FT	CARBOHYD	1919	1919		POTENTIAL.
FT	CONFLICT	713	713		N -> M (IN REF. 2).
FT	CONFLICT	734	734		I -> T (IN REF. 2).
FT	CONFLICT	792	792		G -> Q (IN REF. 2).
FT	CONFLICT	1133	1133		E -> F (IN REF. 2).
FT	CONFLICT	1191	1191		I -> L (IN REF. 2).
FT	CONFLICT	1209	1209		R -> F (IN REF. 2).
FT	CONFLICT	1437	1437		C -> G (IN REF. 2).
FT	CONFLICT	1456	1456		F -> R (IN REF. 2).
FT	CONFLICT	1539	1539		F -> R (IN REF. 2).
FT	CONFLICT	1546	1546		Q -> N (IN REF. 2).
SEQ	SEQUENCE	2133 AA;	239304 MW;		07FCA01E CRC32;

Query Match Best Local Similarity 80.9%; Score 6306; DB 1; Length 2133;
Matches 1227; Conservativity 57.5%; Pred. No. 0;
Indels 678; Gaps 3;

Db	301	HRQASLEISPUETTLTAQTFLEMDJGQFLFCHSHHGGHMEAHVRESQAEPPOLRRAD	360
Qy	360	EEADYDDDLDSMDVAVREFDDDNSEFIOIRSVAKKHBTWVHTIAAEEEDMDYAPLVL	419
Db	361	EE-EDYDNLXDSDMDVRLDGDGDPFQIOIRSVAKKHBTWVHTIASAEEEDMDYAPVP	419
Qy	420	APDRSXYKSOYLNNGPQIRGKXKXKXVAFMAYPDTEFKTEALQHESGILGPLLYGEVDT	479
Db	420	SPSDRSXYKSLYLNNGSPQIRGKXKXKARFAVAYTDFTEFKTKRAIPRESGILGPLLYGEVDT	479
Qy	480	LLIEFKNOASRPNIYVPHGTTDVRPLXSRRLPRGVXHLKDFPILEGELFKYMYTTVYDGG	539
Db	480	LLIEFKKXASRPNIYPHGTTDVSALHPGRLKGWHLKMDMPILPGEETFKYMYTTVYDGG	539
Qy	540	PTKSDPCLLTRYSSFPVNMRLDASGLIGPLLCYKESYDQNGNOJMSKRNVILFVSFD	599
Db	540	PTKSDPCLLTRYSSSINLEKDLASGLIGPLLCYKESYDQNGNOMSDKRVILFVSFD	599
Qy	600	ENRSWYLTENIORFLPNPAGVQLEDEPFQASNMHSINGYVDSQLSVCLHEVAYWYL	659
Db	600	ENOSWYLAENIORFLPNPDGLQPODEPFQASNMHSINGYVDSQLSVCLHEVAYWYL	659
Qy	660	SIGAQOTELSVFSGYFKFKMYEDTTLTFPSGSGTYVMSMENGLWTGCHNSDFNR	719
Db	660	SVGQOTPLSVFSGYFKFKMYEDTTLTFPSGSGTYVMSMENGLWVWGCHNSDLNR	719
Qy	720	GMTALLKVVSCDKNTGDYDYSDYEDIASYLLSKNNAIEBRFSQON-----	765
Lb	720	GMTALLKVVSCDKNTGDYDYSDYEDIASYLLSKNNAIEBRFSQON-----	779
Qy	765	-----	765
Db	780	SPEEDVELDPOSGERTQALEELSVPSGDGSMJLGNPAPHGSSSDJLEARNADYDLPG	839
Qy	765	-----	765
Db	840	ARENTPASAAARLRPLHLHSAERYLTPPEPEKELKLDKSMSSSDJLKTSPITSDLS	899
Qy	765	-----	765
Db	900	AETERTSLSGPHQPVNFRSOLGAIVLGKNSHF IGAGVPLGSTEEDHSSIGENVSPVE	959
Qy	765	-----	765
Db	960	SDGIFEKERAHPASLTKDYLEFKVNISLYTKTKARVYLTKTRKIHIDDAALLTENRASA	1019
Qy	765	-----	765
Db	1080	AVEEELSKGEMMLPNSELTFLTNSADVQGNTHSQKKSEEMERREKQLQKXVDLPQ	1139
Qy	765	-----	765
Db	1140	VYTATGKTNFLRNIFHSTEPSVEGDFGSHAFVPODSRLNDSERAETHIAHFSARE	1199
Qy	765	-----	765
Db	1200	EAPLEAGNRTGPPRSAYVRVXQSLQJRLPLLEIKERGVVLNATSTRNSESPILO	1259
Qy	765	-----	765
Db	1260	GAKRNNLSLPFLTLEMAGOGKISALGKSAAGLAGLKEKAVLSAGLSEASGKAEFLP	1319
Qy	765	-----	765
Db	1320	KVRVHREDLPLQKTSJNWSCAHGLQGEIFLQKTRGPVNLKVNRRPGRTPSKLLGPPMRE	1379
Qy	765	-----	765

Db	1380	WESLEKSPKSTALRTKDILSLPDRHESNHSIAANNEGOAETQREAAWTKOGGPGRLCAP	1439
Qy	765	-PVLKRRHOREIRFTTLOSQDEIEDVDPTISVEMKKKEPDIDEDENOSPSPFOKTRHY	823
Db	1440	KPPLRLRHQNDISLPTQFPEEDMDYDILFSTETGGEFDIDGEBNDPPSPOKRTRHY	1499
Qy	824	FIAAVERIAMDYGMSSSPHYLRNRAQSGSVPOFKVYFOEFTDGSFTOPLYRGELNEHGL	883
Db	1500	FIAAVEQUMDYGMSESPRALNRNQAQGEVPRFKVYREFADGSPFTQSYSGELKHLGL	1559
Qy	884	LGPIRAVEEDNIWYTFRRNQAAPRYFFYSLSLISYEDDQROGAEPRKNVYKPREKTYFWK	943
Db	1560	LGPIRAVEEDNIWYTFRRNQAAPRYFFYSLSLISYEDDQROGAEPRKNVYKPREKTYFWK	1619
Qy	944	VQHMALPTKDEPQCAKAAYSVDVLEKDVHSGILGVLCHNTLNPHAGROYVQOEAL	1003
Db	1620	VQHMALPTDEDFCKAAWYFSDVLEKDVHSGILGVLCHNTLNMAHGRVYVQOEAL	1679
Qy	1004	FFTTFDETKSWYFTENNERNCRAPACNIQMEDPTFKENYFFHAINGYIMDTPLGLVMAODQ	1063
Db	1680	FFTTFDETKSWYFTENNERNCRAPACNIQMEDPTFKENYFFHAINGYIMDTPLGLVMAQNO	1739
Qy	1064	RIRWYLLSMGSGNENIHSHFSGHVFTYRKKEEKYKALVYLVGFVEYEMLPKAGIWRV	1123
Db	1740	RIRWYLLSMGSGNENIHSHFSGHVFTYRKKEEKYKAAVYMLYGVFEVEMLPSPKGIWRI	1799
Qy	1124	ECLIGELHAGMSTLFLYVNSKCOPTLGMASGIRFOFTASGOYGOMAPKLARLHYSGS	1183
Db	1800	ECLIGELHAGMSTLFLYVNSKCOAPLGMASGIRFOFTASGOYGOMAPKLARLHYSGS	1859
Qy	1184	INANSTKEPESWIKVDLALASMIHGIRKTOGARCKFSSLSITSOFLIMSJLDGKKWOTYBGN	1243
Db	1860	INANSTKDPHSWIKVDLALAMIHGIMTQOARGKFSLSYISOFILIMSJLDGNNQSYBGN	1919
Qy	1244	STGLMPLFFGVNDSSGKIKHIEFNPPIIARIKLPHPHYSIRSTLMEIAGCDLNSCMPL	1303
Db	1920	STGLMPLFFGVNDSSGKIKHIEFNPPIVARIKLPHPHYSIRSTLMEIAGCDLNSCMPL	1979
Qy	1304	GMESKALSDAQITASSYFTMMFATWSPSKRRLHLOGRSNAAMPVONNPREMLOYVFOKTM	1363
Db	1980	GMOKALSDQSITASSHLSNIEFTWSPSQRLHLOGRTNAMPRPSSAEMLOVDQKTV	2039
Qy	1364	KVTGVTQGVKSLTISMVYEFLLSISSODGHOWTLEFFQNGKRVYFQGNDSFTPVVNSLD	1423
Db	2040	KVTGITQGVKSLTISMVYKVEFLVSSQDGRMTLEFDQGHKRVYFQGNDSSTPVVNSLD	2099
Qy	1424	PPLLTRYLRIHPQSWHOIALRMEVLCGEADLY	1457
Db	2100	PPLFTYRLRIHPQSWHOIALRLEVLCGEADLY	2133
RESULT 3			
F8A_MOUSE			
ID	F8A_MOUSE	STANDARD;	PRT; 2319 AA.
AC	006194;		
DT	01-JUN-1994 (REL. 29, CREATED)		
DT	01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)		
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)		
DE	COAGULATION FACTOR VIII PRECURSOR (PROCOAGULANT COMPONENT).		
GN	CF8 OR F8C.		
OS	MUS MUSCULUS (MOUSE).		
OC	EDUAROTA; METAQO; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;		
OC	RODENTIA; SCUROGNATHI; MURIDAE; MORINAE; MUS.		
RA	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-C57BL/6 X CBA; TISSUE-LIVER;		
RX	MEDLINE; 93300511.		
RA	ELDER B., LAKICH D., GITSCHER J.;		
RT	"Sequence of the murine factor VIII cDNA."		
RL	GENOMICS 16:374-379(1993).		
CC	-1- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS		
CC	AS A COFACTOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE		
CC	ACTIVATED FORM, FACTOR XA.		

[illegible]

```

QY 765 ----- 765
Db 839 DHSNPAIDSNBGPCKVYOLRPSBHSSEKIVFTPOPGLDRSKNSLETTIEVKMKKGLQV 898
QY 765 ----- 765
Db 899 SSLPSNLTMTTLLSDNLKATEKTDSSGFPDMPVHSSSKSLSTAFGRKAYSISVSHVPLN 958
QY 765 ----- 765
Db 959 ASEENSNDNILDSTLMYSQESLPRDNLISLNDRLLRKREFGIALTKDNTLFRKDNVSL 1018
QY 765 ----- 765
Db 1019 MKTKNTYHSTTNEKLTESPTLSIENSTTDLODAILKYNSEIOETALIHGCTLIGKNST 1078
QY 765 ----- 765
Db 1079 YLRNLHMLNRTTSTKNKIDIFHRKDEDPIDPEENTIMPFCKMLFLESSENNFKKTNGNS 1138
QY 765 ----- 765
Db 1139 LNSEOHSKPOLVYLMFKKYNKNSFLSEKNKVTVEODGTKNIGLKDMAPPHNMSIFLT 1198
QY 765 ----- 765
Db 1199 TLSNVHENGHRNQEKNIOEBIEKEALIEKVVLPQVHEATGSKNFLKDLILIGTRQNTSL 1258
QY 765 ----- 765
Db 1259 YEYHVPVLQNTISINNSTNTVOIHMEHFRRKRRKDETNSEGLVNTREMYKNYPSQKNIT 1318
QY 765 ----- 765
Db 1319 TQSRKALGOFRLSTQWLKNTINGCQCIKOJDSKEMKFKITKSSLSDSYIKSTQTN 1378
QY 765 ----- 765
Db 1379 SSDSHIVTSAPPIDLKRSPQKNFHVQVASSYIDFTKSSRIOESNNFLKTKINP 1438
QY 765 ----- 765
Db 1439 SLALPMNMFIDQKFTSPGKSNNTSVYKKRENTIIFLPTLPRESGKIELLPQVSIQEE 1498
QY 765 ----- 765
Db 1499 EILPETHSGSPGHLNLAKEVFLAQIQGPTKWNKAKRHGESIKGTSSKTRSKLNNH 1558
QY 765 ----- 765
Db 1559 AMDHYAAQIPKDMWKSKEKSPETISIKOEDTILSLRPHGNSHIGANEKONMPQRETTW 1618
QY 765 ----- 765
Db 1619 VKQGTQRTCSQIIPVLRHQRREL--SAFQSQEATDIDDATEITIE-TIEDPDYISEDIKQ 1675
QY 812 SPRSQKTRHYFAAVERLMDYGSPPHYLRNAGSGSVPOFKVYVQFETDSSFTOP 871
Db 1676 GPRSQKTRHYFAAVERLMDYGSPPHYLRNAGSGSVPOFKVYVQFETDSSFTOP 1734
QY 872 LYRGLNLHLGLGPIYIAEVEDNIMVFRNOSAPRYSFYSSLSIYEDDQAGAPRKNF 931
Db 1735 LYRGLNLHLGLGPIYIAEVEDNIMVFRNOSAPRYSFYSSLSIYEDDQAGAPRKNF 1793
QY 932 VKPNETKTYFMVQHHMAPTKDEPCKAMAYSDVDLEKDVSGILGFLVCHTNTLPA 991
Db 1794 VKPNETKTYFMVQHHMAPTKDEPCKAMAYSDVDLEKDVSGILGFLVCHTNTLPA 1853
QY 992 HGRQVTVQVEFALFTIPDETQSWYFTENNERCRAPCNIONEDPTFKENYRHAINGYM 1051
Db 1854 HGRQVTVQVEFALFTIPDETQSWYFTENNERCRAPCNIONEDPTFKENYRHAINGYM 1913
QY 1052 DTLPLVMAQODORIRMYLLSHMSNENHSHFSGHVFTYRKKEEKMALYNLYPQVETV 1111

```

```

Db 1914 DTLPLVMAQODORIRMYLLSHMSNENHSHFSGHVFTYRKKEEKMAVNLVPGVETL 1973
QY 1112 EMLPSKAGINRVECLLIGHLHAGMSTFLVYSNKQCPPLGNAASHIRDFQTASGQVGM 1171
Db 1974 EMIPSRAGINRVECLLIGHLHAGMSTFLVYSNKQCPPLGNAASHIRDFQTASGQVGM 2033
QY 1172 APKLARLHSGSINAMSTKEPFSWIKVDLAPMIIHGKITGAGAROKESSLYISQFIWYS 1231
Db 2034 APKLARLHSGSINAMSTKEPFSWIKVDLAPMIIHGKITGAGAROKESSLYISQFIWYS 2093
QY 1232 LDGKKMOTYRGNSTGTLVWFGNVDSSGIRKNTINPPIATYILRPHVYSIRSTRLMEL 1291
Db 2094 LDGKKMOTYRGNSTGTLVWFGNVDSSGIRKNTINPPIATYILRPHVYSIRSTRLMEL 2153
QY 1292 MGDNLNSCMBLGMESKASIDAOITASSYFTNMFATSPSKARLHLOGRSNAMPQVNP 1351
Db 2154 MGDNLNSCMBLGMESKASIDAOITASSYFTNMFATSPSKARLHLOGRSNAMPQVNP 2213
QY 1352 KEMLOVDFQKTKMYTGYTQGVKSLTSMYKEFLISSQDGHWTFLFQNGKYVFOGN 1411
Db 2214 KEMLOVDFQKTKMYTGYTQGVKSLTSMYKEFLISSQDGHWTFLFQNGKYVFOGN 2273
QY 1412 QDSFTPVNLSLDPPLRFLYLRHPOSWVHOALNMEVYGCENODLY 1457
Db 2274 QDSFTPVNLSLDPPLRFLYLRHPOSWVHOALNMEVYGCENODLY 2319

RESULT 4
FAS_HUMAN STANDARD; PRT; 2224 AA.
ID P12259; Q14285;
AC 01-OCT-1989 (REL. 12, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE COAGULATION FACTOR V PRECURSOR (ACTIVATED PROTEIN C COFACTOR).
GN F5.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX CRPE L.D., MOORE K.D., KANE W.H.;
RT "Structure of the gene for human coagulation factor V.";
RL BIOCHEMISTRY 31:3777-3785(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87260886.
RA JENNY R.D., PITTMAN D.D., TOOLE J.J., KRIZ R.W., ALDAPE R.A.,
RA HENICK R.M., KAUFMAN R.J., MANN K.G.;
RT "Complete cDNA and derived amino acid sequence of human factor V.";
RL PROC. NATL. ACAD. SCI. U.S.A. 84:4846-4850(1987).
RN [3]
RP SEQUENCE OF 1-1600 FROM N.A.
RX MEDLINE; 88107360.
RA KANE W.H., ICHINOSE A., HAGEN F.S., DAVIE E.W.;
RT "Cloning of cDNAs coding for the heavy chain region and connecting
RT region of human factor V, a blood coagulation factor with four types
RT of internal repeats.";
RL BIOCHEMISTRY 26:6508-6514(1987).
RN [4]
RP SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
RX MEDLINE; 86313665.
RA KANE W.H., DAVIE E.W.;
RT "Cloning of a cDNA coding for human factor V, a blood coagulation
RT factor homologous to factor VIII and ceruloplasmin.";
RL PROC. NATL. ACAD. SCI. U.S.A. 83:6800-6804(1986).
RN [5]
RP PARTIAL SEQUENCE FROM N.A.
RX TISSUE-FIBROBLAST;
RX MEDLINE; 93203619.
RA SHEN N.L.L., FAN S.-T., PYATT J., GRAFF R., LAPOLLA R.J.,

```

RA EDGINGTON T.S.:
 RT "The serine protease cofactor factor V is synthesized by
 RT lymphocytes."
 RL J. IMMUNOL. 150:2992-3001(1993).
 RN [6]
 RP VARIANT APCR GLN-534.
 RX MEDLINE: 94217810.
 RA BERTINA R.M., KOELEMAN B.P.C., KOSTER T., ROSENDAL F.R.,
 RA DIEREN R.J., DE RONDE H., VAN DER VELDEN P.A., REITSMA P.H.:
 RT "Mutation in blood coagulation factor V associated with resistance to
 RT activated protein C."
 RL NATURE 369:64-67(1994).
 CC -I- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
 CC WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
 CC -I- SUBUNIT: FACTOR VA IS COMPOSED OF AN HEAVY CHAIN AND OF A LIGHT
 CC CHAIN NONCOVALENTLY BOUND. THE INTERACTION BETWEEN THE TWO CHAINS
 CC IS CALCIUM-DEPENDENT.
 CC -I- DOMAIN: DOMAIN B CONTAINS 35 X 9 AA TANDEM REPEATS, AND 2 X 17 AA
 CC REPEATS.
 CC -I- PTM: THROMBIN ACTIVATES FACTOR V PROTEOLYTICALLY TO THE ACTIVE
 CC COFACTOR, FACTOR V(A) (FORMATION OF A HEAVY CHAIN AT THE N-
 CC TERMINUS AND A LIGHT CHAIN AT THE C-TERMINUS).
 CC -I- DISEASE: OMEN PARAHOMOPHILIA, AN HEMORRHAGIC DIATHESIS, IS DUE
 CC TO A DEFICIENCY OF FACTOR V. OTHER DEFECTS IN F5 RESULTS IN A
 CC FORM OF THROMBOPHILIA KNOWN AS APC RESISTANCE (APCR). THE APCR
 CC MUTATION IS FOUND IN ABOUT 5% OF THE POPULATION WHICH SUGGEST THAT
 CC A SLIGHT THROMBOTIC TENDENCY MAY CONFER SOME ADVANTAGE IN FETAL
 CC IMPLANTATION.
 CC -I- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
 CC 2 PLASTOCYANIN-LIKE REPEATS.
 CC -I- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -I- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: L32779; G488110; -
 DR EMBL: L32755; G488110; JOINED.
 DR EMBL: L32756; G488110; JOINED.
 DR EMBL: L32757; G488110; JOINED.
 DR EMBL: L32758; G488110; JOINED.
 DR EMBL: L32759; G488110; JOINED.
 DR EMBL: L32760; G488110; JOINED.
 DR EMBL: L32761; G488110; JOINED.
 DR EMBL: L32762; G488110; JOINED.
 DR EMBL: L32763; G488110; JOINED.
 DR EMBL: L32764; G488110; JOINED.
 DR EMBL: L32765; G488110; JOINED.
 DR EMBL: L32766; G488110; JOINED.
 DR EMBL: L32767; G488110; JOINED.
 DR EMBL: L32768; G488110; JOINED.
 DR EMBL: L32769; G488110; JOINED.
 DR EMBL: L32770; G488110; JOINED.
 DR EMBL: L32771; G488110; JOINED.
 DR EMBL: L32772; G488110; JOINED.
 DR EMBL: L32773; G488110; JOINED.
 DR EMBL: L32774; G488110; JOINED.
 DR EMBL: L32775; G488110; JOINED.
 DR EMBL: L32776; G488110; JOINED.
 DR EMBL: L32777; G488110; JOINED.
 DR EMBL: L32778; G488110; JOINED.
 DR EMBL: M16967; G182412; -
 DR EMBL: M14335; G182798; -
 DR PIR: A25897; A25897.
 DR PIR: A28028; A28028.
 DR MIM: L34400; -
 DR MIM: L68055; -
 DR MIM: L27310; -

DR MIM: 227400; -
 DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 2.
 DR PROSITE: PS01285; FA58C.1; 2.
 DR PROSITE: PS01286; FA58C.2; 2.
 DR PFAM: PF00394; Cu-oxidase; 3.
 DR PFAM: PF00734; F5_F8_Type_C; 2.
 DR HSSP: P00450; IKCW.
 DR BLOOD COAGULATION; PLASMA; GLYCOPROTEIN; CALCIUM; SIGNAL; ZMOGEN;
 KW REPEAT: POLYMORPHISM; DISEASE MUTATION; THROMBOPHILIA.
 FT SIGNAL
 FT CHAIN 1 28
 FT CHAIN 29 2224
 FT PEPTIDE 738 1573
 FT CHAIN 1574 2224
 FT DOMAIN 30 329
 FT DOMAIN 30 193
 FT DOMAIN 203 329
 FT DOMAIN 348 684
 FT DOMAIN 348 526
 FT DOMAIN 536 684
 FT DOMAIN 692 1573
 FT DOMAIN 895 928
 FT DOMAIN 895 911
 FT REPEAT 912 928
 FT SIMILAR 1135 1148
 FT DOMAIN 1185 1463
 FT REPEAT 1185 1193
 FT REPEAT 1194 1202
 FT REPEAT 1203 1211
 FT REPEAT 1212 1220
 FT REPEAT 1221 1229
 FT REPEAT 1230 1238
 FT REPEAT 1239 1247
 FT REPEAT 1248 1256
 FT REPEAT 1257 1265
 FT REPEAT 1266 1274
 FT REPEAT 1275 1283
 FT REPEAT 1284 1292
 FT REPEAT 1293 1301
 FT REPEAT 1302 1310
 FT REPEAT 1311 1319
 FT REPEAT 1320 1328
 FT REPEAT 1329 1337
 FT REPEAT 1338 1346
 FT REPEAT 1347 1355
 FT REPEAT 1356 1364
 FT REPEAT 1365 1373
 FT REPEAT 1374 1382
 FT REPEAT 1383 1391
 FT REPEAT 1392 1400
 FT REPEAT 1401 1409
 FT REPEAT 1410 1418
 FT REPEAT 1419 1427
 FT REPEAT 1428 1436
 FT REPEAT 1437 1445
 FT REPEAT 1446 1454
 FT REPEAT 1455 1463
 FT REPEAT 1464 1472
 FT REPEAT 1473 1481
 FT REPEAT 1482 1490
 FT REPEAT 1493 1501
 FT DOMAIN 1578 1907
 FT DOMAIN 1578 1751
 FT DOMAIN 1761 1907
 FT DOMAIN 1907 2061
 FT DOMAIN 2066 2221
 FT SITE 737 738
 FT SITE 1046 1047
 FT SITE 1573 1574
 FT DISULFID 167 193
 FT DISULFID 500 526
 FT DISULFID 1725 1751
 COAGULATION FACTOR V.
 HEAVY CHAIN.
 ACTIVATION PEPTIDE (CONNECTING REGION).
 LIGHT CHAIN.
 F5/8 TYPE A 1.
 PLASTOCYANIN-LIKE 1.
 PLASTOCYANIN-LIKE 2.
 F5/8 TYPE A 2.
 PLASTOCYANIN-LIKE 3.
 PLASTOCYANIN-LIKE 4.
 B.
 2 X 17 AA TANDEM REPEATS.
 1.
 2. 14 AA REPEATS IN BOVINE FAS.
 35 X 9 AA TANDEM REPEATS OF [TNP]-L-S-P-
 D-L-S-Q-T (APPROXIMATE).
 1.
 2.
 3.
 4.
 5.
 6.
 7.
 8.
 9.
 10.
 11.
 12.
 13.
 14.
 15.
 16.
 17.
 18.
 19.
 20.
 21.
 22.
 23.
 24.
 25.
 26.
 27.
 28.
 29.
 30.
 31.
 32.
 33.
 34.
 35.
 F5/8 TYPE A 3.
 PLASTOCYANIN-LIKE 5.
 PLASTOCYANIN-LIKE 6.
 F5/8 TYPE C 1.
 F5/8 TYPE C 2.
 CLEAVAGE (BY THROMBIN).
 CLEAVAGE (BY THROMBIN).
 CLEAVAGE (BY THROMBIN).
 PROBABLE.
 PROBABLE.
 PROBABLE.

FT DISULFID 1907 2061 BY SIMILARITY.
FT DISULFID 2066 2221 BY SIMILARITY.
FT CARBOHYD 51 51 POTENTIAL.
FT CARBOHYD 55 55 POTENTIAL.
FT CARBOHYD 239 239 POTENTIAL.
FT CARBOHYD 297 297 POTENTIAL.
FT CARBOHYD 382 382 POTENTIAL.
FT CARBOHYD 460 460 POTENTIAL.
FT CARBOHYD 468 468 POTENTIAL.
FT CARBOHYD 554 554 POTENTIAL.
FT CARBOHYD 741 741 POTENTIAL.

Query Match 30.7% Score 2390.5; DB 1; Length 2224;
Best Local Similarity 26.18; Pred. No. 2.6e-146;
Matches 593; Conservative 276; Mismatches 464; Indels 923; Gaps 35;

22 RRYIAGAVELSMYQSDGLGELPVDAREPPVPRKSPFNTSVV-YKTLFEVETDHLFNI 80
32 RQFYVAAGISMSYRPE-----PNNSSLNLSVTGFKIYREYEPY-FKK 75
81 AKRPPPMGLGPTIQAEVYDVVITLKNMASHVSLHAGVSYWKASEGAYDQTSQR 140
76 EKPGSTISGLGPTLYAEVDGIKVFHKNKADKPLSIHQIGIRYSKLSGASGLDHTFPA 135
141 EKEDDKYFPGSGHTYVQVLEKNGPMASDPLCTYSYLSHDLVKDLNSGLIGALLYCRE 200
136 EKMDAAVAPGHEHYEWSISEDSGPTHDDPCLTHIYSHENLLEDNPSGLIGLILCKK 195
201 GSLAKEKTO-TLHK-FILLFAVDEGKSMHSETKNSLMQDRDASARAPMHTVNGVYN 258
196 GTLEGGTQKTFDKQIYLLFAVFEDESKMSQSSS-----LMYVNGVYN 239
259 RSLPGLIGCHRRKSVYHVGMTTPEVHSIFLEGHFTLVRRNQASLEISPTFLAQT 318
240 GTMPDITVACADHISMHLGMSGPELFSIHFGQVLEQNHKYSATITLVASSTANMT 299
319 LMDGQTLFCHISSHODGMEAYVYKVDSCPEBOLRKKNNEEDVDLDTSEMDVYR 378
300 VGPBGKMISSLTPKHLQAGQAYIDIKNCPRKTRNLKTRRE----- 343
379 FDDNSPSFIQIRSVAKHPRKTVHYIAEEDMDYAPLVAPDRSKSYQLNNGPQRI 438
343 -----QRHMKRMEYFIAEEVINDYAPVIAANDKTKRSHLDNFSQI 387
439 GRKYKVRFYMAVYDETEKTRAI---OHSGLIGLPLYGEVDTLLIFKNQASRPYNY 495
388 GKHYKKVWYQYEDSEF-TKHTVNPNMKEGDIIGPTIRAQVRDLKIYFKMMASRPYSIY 446
496 PHGIT-----DVRPLYSRLLPKGVKHLKDPILPGEIYFKKWTYVEDGPTKSDPRLCT 549
447 PHGTFSPYEDEVNSFT---SGRNNTMIRAVOPGETYTYKMWILFEDEPTENDACILT 502
550 RYSSSFVNMEDLASGLIGLILCYKESVDORGNOIMSDKNNVILFSVFDENRSMYLTEN 609
503 RPYSDVDIMDIASGLIGLILCKSRSLDRGIORADIDQAVFAFEDNKKWYLEDN 562
610 IQRLFPNAGVLEDPFEOASINMHSINGYFDSL-QLSVCLHEVAYVYISISAGQDEL 668
563 INKCEINDEVKRDQPKYESINIMSTINGYPESTITLGFCEFDYQVQHFCVSYQONIL 622
669 SVFSGYTFKKKMYEDTLTFPPSGETVEMSMENPGMLILGCHNS----- 715
623 TIHGTGHSFIGKRHEDTLTFPMKGSVSYVTMDNVGTWMLTSMNSPRSKRLKRFDV 682
715 -----DFPNR----- 720
683 KCIPIDDSDSEIPEPPESTYMATRKMDRLPEDEESDADYDQNKLAALGIRSFNS 742
720 -----GMTAL----- 725
743 SLNDEEEFNLTALALENGTEFVSSNTDIIYGSNYSPPSNISKFTVNNLAPQKAPSHQ 802
725 -----LKVSSCDKNTGDIYEDSYED----- 745

DB 803 ATTAGSPILRLHIGKNSVLSNSTAEHSPISYEDPIEDLPDVTGIRLLSGLAGEFKSOEH 862
745 ----- 745
863 AKHKGPVERDQAAKHFRSMKLLAHKVGRLSODTGSFGMRPWEDLPQDITGSPSRMR 922
745 -----ISAYLSKNN----- 756
923 PMKDPSPDLLLKQSNSSKILVGRWHLASEKSEYIIOPTDEDTAVANNMLISQNASRAW 982
756 ----- 756
983 GESTPLANKGKSGHPPKPRVRKHSLOVYQDGGKSRLLKSQTLITRKKKKEKHTHHAP 1042
756 IEPSSF-----SQ 763
1043 LSPRTFHLRSEAYNPFSEERLKHSLVHKSNETSLEPTDLNQTLPMSDFGMSLPDHNQ 1102
764 N----- 765
1103 NSSNDTQASCPPGLYQVTPPEEHYQTFPIODPDQMHSTSDPSHRSSPELSEMLEYDRS 1162
765 -----PVLKRH----- 772
1163 HKSFPTDISQMSPSSEHEWQVYISPDLSQVTLSPELSQTNLSBDLSHTLSPELLQRL 1222
772 ----- 772
1223 SPALGQPIPSDLSTHTLSPDLSTHTLSPDLSTHTLSPDLSTHTLSPDLSTHTLSPDL 1282
772 ----- 772
1283 TTLSLDSQTNLSPELSHMTLSPELSQTNLSPALGQMPISPDLSHTLSPDLSTHTLSP 1342
772 -----QRETRTTLSDQEI----- 788
1343 LSQTNLSPALGQMPSPDPSHTLSPDLSTHTLSPDLSTHTLSPDLSTHTLSPDLSTHT 1402
788 ----- 788
1403 TPDLQNTLSPDLGPTLSPNFGMSLSPDLSTHTLSPDLSTHTLSPDLSTHTLSPDL 1462
788 -----DYDT 792
1463 IFYSSSSQSLLOEFNESEFPYDLPQMPSPSPPTLNDTFLSKFNPVLVGLSKDCTDY 1522
793 ISVEMKE-----DEDIYE-----DENSP-----RSFOKTRHY 823
1523 IEIIPKEEVOSSEDDYAEIDYVPPDYDKTDVRTNINSSRDPONIAAMYLRSNNGNRNY 1582
824 FIAVERLMDYGMSSSPHYLRN--AQSGSVQ---FKKYVFOFTGSGTQPLRYGELN 878
1583 YIAAEISMDY---SEFVORETIDIEDSDIPETTYTKKVFYKYLDTSTFKRDPREYE 1638
879 EHLGLIPYIAEVEDNINMFRNOASRPYSFYSSLSYE-----EDOROGAEPRKN 930
1639 EHLGLIPYIAEVEDNINMFRNOASRPYSFYSSLSYE-----EDOROGAEPRKN 930
931 FVKNETKTYFWKQVHMAAPTKDEFCKANAYVSDVLEKDVHSGLIGLILVCHTNTLP 990
1699 AVQNSSTYVWMAHTEKSGEPPSACRAVAYSAVPEKDIHSGLIGLILVCHTNTLP 1758
991 AHGRQVTVQEFALFEFTLEDETCKSVYFTENNERBNCRAPCNOMEDPTKENVRFPAINGYI 1050
1759 DSNRPVDMREFVLLFMFDEKSKSVYTEKKSRSNR-----LISSEKKSHHEHAINGMI 1812
1051 MDTLPLGLVMAQDQIRIRYLLSMGSENENIHSIFSGHYFTVRRKEEYKMALYNYIPVGET 1110
1813 Y-SLPGIKTAEQEWVRLHLNLINIGSODIHVYHFGQTLLENGNQHQGLGVWPLPGSFKT 1871
1111 VEMLPKRAGITRVCCLGELHAGMSTLFLVYSKKCTPPLGASGHTRDQITASGQYQ 1170

```

Db 1872 LEMKSRKPGWMLNTEVEGNQAGMOTPELIMDRCDRMPMGISTGIIISDQIKASEFLGY 1931
Oy 1171 WAPKIALRHYSGSINAMSTKE---PFS---WIKVDLLPMIHHGKTKGAGKQKSSLYIS 1224
Db 1932 WEPRLARLNGNSYNAWEKLAFAASKPQVDMOREVITGOTGQAHYLYKSCOTT 1991
Oy 1225 OFIIMYSLDGKKWQTYRGSGTGLTAVFPGVNDSSGKININIEPPIARIYIRLPHYSIR 1284
Db 1992 EPIYVASSNOIMWQIFKGNSTFNWYFNGNSDASTIKENQPPPIVAYIRISPRRAVN 2051
Oy 1285 STLRMLMGCDLNSCSMPLMESKASISDAQITASSYFTINMEAT-WSPSKARLHLOGRSNA 1343
Db 2052 PTLRLDLGCEVNGCSTPLGMEKNGKIENKQITASSFKKSMWDYEPFRARLMAQGRVNA 2111
Oy 1344 WEPQVNNKREMLQVDFQTKMKTGYTTCGVKSLTSMVYKEPLISSODGHQWTLFQNG 1403
Db 2112 WQAKANNKQWLEIDLKTKITAITGCKSLSEMYKSTYTHYSEGVEMKPYRLKS 2171
Oy 1404 KY--KVFQGNDSFTPVVNSLDPPLTLRLYLRHPQSWVHQLRMVEVLGCEADQLY 1457
Db 2172 SNAWKIFEGNTKGVKHFNFPPILSRIRYIPKTNOSITLRLDLGCG---DIY 2224

RESULT 5
FAS_BOVIN STANDARD: PRT: 2211 AA.
AC 028107: 028108:
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE COAGULATION FACTOR V PRECURSOR (ACTIVATED PROTEIN C COFACTOR).
GN F5.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDEA; BOVIDAE; BOVINAE; BOS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE: 92147638.
RA GUINTO E.R., ESMON C.T., MANN K.G., MACGILLIVRAY R.T.;
RT "THE COMPLETE CDNA SEQUENCE OF BOVINE COAGULATION FACTOR V.";
RL J. BIOL. CHEM. 267:2971-2978(1992).
CC -1- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
CC WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
CC -1- SUBUNIT: FACTOR VA IS COMPOSED OF AN HEAVY CHAIN AND OF A LIGHT
CC CHAIN NONCOVALENTLY BOUND. THE INTERACTION BETWEEN THE TWO CHAINS
CC IS CALCIUM-DEPENDENT.
CC -1- DOMAIN: DOMAIN B CONTAINS 29.5 X 9 AA TANDEM REPEATS, AND 2 X 17
CC AA REPEATS.
CC -1- PTM: THROMBIN ACTIVATES FACTOR V PROTEOLYTICALLY TO THE ACTIVE
CC COFACTOR, FACTOR V(A) (FORMATION OF A HEAVY CHAIN AT THE N-
CC TERMINUS AND A LIGHT CHAIN AT THE C-TERMINUS).
CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC 2 PLASTOCYANIN-LIKE REPEATS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M81440: G163038: -
CC EMBL: M81441: G163040: -
DR PROSITE: PS00079; MULTICOPPER_OXIDASEL. 2.
DR PROSITE: PS01285; FAS5C_1; 2.
DR PROSITE: PS01286; FAS5C_2; 2.
DR PRAM: PF00394; Cu-oxidase; 3.
DR PRAM: PF00754; F5_F8_type_C; 2.
DR HSSP: P00450; 1KCW.

KW BLOOD COAGULATION; PLASMA; GLYCOPROTEIN; CALCIUM; SIGNAL; ZYMOGEN;
REPEAT.
KM REPEAT.
FT SIGNAL. 1 28
FT CHAIN. 29 2211
FT CHAIN. 29 741
FT CHAIN. 742 1564
FT PEPTIDE
FT CHAIN 1565 2211
FT DOMAIN 30 327
FT DOMAIN 30 193
FT DOMAIN 203 327
FT DOMAIN 348 686
FT DOMAIN 348 525
FT DOMAIN 535 686
FT DOMAIN 696 1564
FT SIMILAR 1124 1151
FT DOMAIN 1124 1137
FT REPEAT 1124 1137
FT REPEAT 1138 1151
FT DOMAIN 1188 1453
FT REPEAT 1188 1196
FT REPEAT 1197 1205
FT REPEAT 1206 1214
FT REPEAT 1215 1223
FT REPEAT 1224 1232
FT REPEAT 1233 1241
FT REPEAT 1242 1250
FT REPEAT 1251 1259
FT REPEAT 1260 1268
FT REPEAT 1269 1277
FT REPEAT 1278 1286
FT REPEAT 1287 1295
FT REPEAT 1296 1304
FT REPEAT 1305 1313
FT REPEAT 1314 1322
FT REPEAT 1323 1331
FT REPEAT 1332 1340
FT REPEAT 1341 1349
FT REPEAT 1350 1358
FT REPEAT 1359 1367
FT REPEAT 1368 1376
FT REPEAT 1377 1385
FT REPEAT 1386 1394
FT REPEAT 1395 1403
FT REPEAT 1404 1412
FT REPEAT 1413 1421
FT REPEAT 1422 1430
FT REPEAT 1431 1439
FT REPEAT 1440 1444
FT REPEAT 1445 1453
FT DOMAIN 1459 1890
FT DOMAIN 1569 1738
FT DOMAIN 1748 1890
FT DOMAIN 1894 2048
FT DOMAIN 2053 2208
FT SITE 741 742
FT SITE 1034 1035
FT SITE 1564 1565
FT DISULFD 167 193
FT DISULFD 499 525
FT DISULFD 1712 1738
FT DISULFD 1894 2048
FT DISULFD 2053 2208
FT CARBOHYD 225 225
FT CARBOHYD 239 239
FT CARBOHYD 297 297
FT CARBOHYD 382 382
FT CARBOHYD 460 460
FT CARBOHYD 553 553
FT CARBOHYD 587 587
FT CARBOHYD 745 745
FT CARBOHYD 756 756

1. POTENTIAL.
2. COAGULATION FACTOR V.
3. HEAVY CHAIN (BY SIMILARITY).
4. ACTIVATION PEPTIDE (CONNECTING REGION)
5. (BY SIMILARITY).
6. LIGHT CHAIN (BY SIMILARITY).
7. F5/8 TYPE A 1.
8. PLASTOCYANIN-LIKE 1.
9. PLASTOCYANIN-LIKE 2.
10. F5/8 TYPE A 2.
11. PLASTOCYANIN-LIKE 3.
12. PLASTOCYANIN-LIKE 4.
13. TO 17 AA REPEATS IN HUMAN FAS.
14. 2 X 14 AA TANDEM REPEATS.
15.
16.
17. 30 X 9 AA TANDEM REPEATS OF [AS]-L-S-P-
18. D-[LP]-[GS]-O-[TE] (APPROXIMATE).
19.
20.
21.
22.
23.
24.
25.
26.
27.
28.
29. (PARTIAL).
30. F5/8 TYPE A 3.
31. PLASTOCYANIN-LIKE 5.
32. PLASTOCYANIN-LIKE 6.
33. F5/8 TYPE C 1.
34. F5/8 TYPE C 2.
35. CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
36. CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
37. PROBABLE.
38. BY SIMILARITY.
39. BY SIMILARITY.
40. POTENTIAL.
41. POTENTIAL.
42. POTENTIAL.
43. POTENTIAL.
44. POTENTIAL.
45. POTENTIAL.
46. POTENTIAL.
47. POTENTIAL.
48. POTENTIAL.
49. POTENTIAL.
50. POTENTIAL.
51. POTENTIAL.
52. POTENTIAL.
53. POTENTIAL.
54. POTENTIAL.
55. POTENTIAL.
56. POTENTIAL.
57. POTENTIAL.
58. POTENTIAL.
59. POTENTIAL.
60. POTENTIAL.
61. POTENTIAL.
62. POTENTIAL.
63. POTENTIAL.
64. POTENTIAL.
65. POTENTIAL.
66. POTENTIAL.
67. POTENTIAL.
68. POTENTIAL.
69. POTENTIAL.
70. POTENTIAL.
71. POTENTIAL.
72. POTENTIAL.
73. POTENTIAL.
74. POTENTIAL.
75. POTENTIAL.
76. POTENTIAL.
77. POTENTIAL.
78. POTENTIAL.
79. POTENTIAL.
80. POTENTIAL.
81. POTENTIAL.
82. POTENTIAL.
83. POTENTIAL.
84. POTENTIAL.
85. POTENTIAL.
86. POTENTIAL.
87. POTENTIAL.
88. POTENTIAL.
89. POTENTIAL.
90. POTENTIAL.
91. POTENTIAL.
92. POTENTIAL.
93. POTENTIAL.
94. POTENTIAL.
95. POTENTIAL.
96. POTENTIAL.
97. POTENTIAL.
98. POTENTIAL.
99. POTENTIAL.
100. POTENTIAL.

```

FT CARBOHYD 774 774 POTENTIAL.
FT CARBOHYD 780 780 POTENTIAL.
FT CARBOHYD 902 902 POTENTIAL.
FT CARBOHYD 952 952 POTENTIAL.
FT CARBOHYD 964 964 POTENTIAL.
FT CARBOHYD 1044 1044 POTENTIAL.
FT CARBOHYD 1053 1053 POTENTIAL.
FT CARBOHYD 1062 1062 POTENTIAL.
FT CARBOHYD 1071 1071 POTENTIAL.
FT CARBOHYD 1078 1078 POTENTIAL.
FT CARBOHYD 1094 1094 POTENTIAL.
FT CARBOHYD 1451 1451 POTENTIAL.
FT CARBOHYD 1490 1490 POTENTIAL.
FT CARBOHYD 1550 1550 POTENTIAL.
FT CARBOHYD 1690 1690 POTENTIAL.
FT CARBOHYD 1839 1839 POTENTIAL.
FT CARBOHYD 1997 1997 POTENTIAL.
FT CARBOHYD 2196 2196 POTENTIAL.
FT VARIANT 587 592 NRTIPA -> T (IN VARIANT 2).
SQ SEQUENCE 2211 AA; 248981 MW; 98017C3C CMC32;

Query Match 30.4%; Score 2370; DB 1; Length 2211;

Best Local Similarity 26.2%; Pred. No. 5.4e-145; Mismatches 487; Indels 924; Gaps 35;

Matches 594; Conservative 265; Mismatches 487; Indels 924; Gaps 35;

QY 22 RRYVLGAVELSDWYQSDGLPVDARPPRPVKSPFNTSVYKKTLEVEFTDHLFNIA 81
DB 32 ROFYVAASIRNMYR-----PESTHL-----SKPFETS--FKIIVREY-EAIVFKE 76

QY 82 KRPRWMMGLPTIOAEVYDVITLKNASHPVSLHANGVYKMASEAEVDDOTSORE 141
DB 77 KQOSTSGGLGTLAEVDIKVHKNAKPLSIHAGIKYKFSKSGASVSDHLLP 136

QY 142 KEDDVFPGSHTYWQVLEKNGPMASDPLCLTYSTLSHVLDVKDLSGLIGALLVCREG 201
DB 137 KMDAVAPQOETYEIMISEHSGPTHDPPLCLHITYSVNLVEDPNSLIGPLICKKG 196

QY 202 SLAKEKOTL--HKFTLFAVDEGKSWHSETKNSLMQDRDAASARAMEKMTVNGYVR 259
DB 197 TLTEGTOKMFQKHVLMFAVEDESKVMTQSS-----LMTVNGYVNG 240

QY 260 SLPLGLGCRKSVYVHVGMTTPEVHSIFLEGHTFLVNNHROASLEIPITLQATLL 319
DB 241 TMDPITVCAHDIHMLIGMSSGPELSIFHNGQVLEQNHKISAITLVASATSTANMTV 300

QY 320 MDLQGLFCHSHQHODMEAYVYKVDSCPEPOLRMKNNNEAEVDLDTLSEMDVYAF 379
DB 301 SPEGKTIASLIPRHQAGMAYIDIKKCAKTRNPK-----LTRDQ----- 344

QY 380 DDNSPSFIQIRSVAKKHKPTWVHYIAEEDMDYAPLVLPDDRSGYSQYLNQFORIG 439
DB 344 -----RRHKRWEYFIAAEVYIMDYAPIIPAMMDKRYSLHLDNSSNRIG 388

QY 440 RYKRYVREMYTDEFTKTR--EAIQHESILGPLYGEVGDILLIFKQASRPYIYTH 497
DB 389 KHYKRVYKQYDSTFKLEDPSSSEGDILGPLIRAYQVRLKIFYKMMASYSIYPH 448

QY 498 GIT-----DVRPLYRRLPKGVKHLKDFPLGELFKYKMTVVDGPTKSPRCL 548
DB 449 GYTFSPYDNEVSSSTSGSNTMRAVR-----PGETTYTKMNLSEDEPTENDAOCL 500

QY 549 TRYSSFNMERDLASGLIGPLLCYKESVDQNGOIMSDKRNVLISFVDENRSWYLE 608
DB 501 TRPYSNVDITRDLASGLIGLLICKSRSLDRGIRADIEQAVFAVFEDEKMSYIED 560

QY 609 NIORELPNAGVQLEDPPEQASINM-----SINGVFDLSQ--LSYCLAEVAYWYTLISG 662
DB 561 NYYKFEENPEKVRDPPKESYENIMSNFTLPAINGVPEIPILGCFPDOTVQMHCECVG 620

QY 663 AOTDFLVSFSGYTFKKNVYEDTLFLPFSGETVMSNENGLWTLGCHNSDFRRKGMF 722
DB 621 TONDILTFHTGSHFYGRHEDTLFLPQGESVTVTMDNGTWMALTMTNSNPRSKLR 680

QY 723 ALLKVSQCDKNTGD-----YEDS-----YEDISAYLL-- 751
DB 681 LRFRAKCIKRNDDDSYEIIYEPGSGTAMTKIKIDSSEIEDENDADSDYODLAILGL 740

QY 751 -----SKNNAIEPRSPQ----- 764
DB 741 RSFRNSLNQEKDELNTALALEKDESEFTPPSANSLDSSSSRSHVSLAKNFAESLK 800

QY 764 -----NPPVLK----- 770
DB 801 TLHLERAPAGSPLHAGLDKNSALNPPMAHSSPYSDPRDHPDLSDVTGVSLLPFGG 860

QY 770 -----RHQR-----EITRTILOSQO-----EET----- 788
DB 861 FKNRPKAKHQRFQVGRGAAKHKFSQTRFPAKTRTRLSQDNSSSRMGPMWEDIPSDLL 920

QY 788 ----- 788
DB 921 LQOKDPYKILNGEWHLVSEKGYEIIQDANENKTVNKLPNSPONDRTWGENIPEKNSHG 980

QY 788 ----- 788
DB 981 KQSGHPTLVYTRKRLQDRODRNRSLKEGLPLITRRKKREKRAYHVPLSPRSHPLR 1040

QY 788 -----DYDD----- 792
DB 1041 GEVNASFSDRRHNSLLHASNETSIDLNQTFPSMNLASLAPDHQTSPPNDTSQOT 1100

QY 792 -----TISYEMKKEDEFDIYDE-----ENQSP----- 814
DB 1101 SSPDLPLVPSPEEHYQIFPQDSPTHTSTAPSNSRSPDHTSTAPSNSRSPDQIP 1160

QY 814 ----- 814
DB 1161 NYDLNRAIPTVSOIFPSLELVQOTATSLDSQPSISPDLGQNALSPDQESLSPDL 1220

QY 814 ----- 814
DB 1221 GQTSLSPLSQESLSPDLGQTALSPDPSQESLSPDLGQTALSPDPSQESLSPDLGQTALS 1280

QY 814 ----- 814
DB 1281 PDGQESLSPDLGQTSLSPLSQESLSPDLGQTALSPDPSQESLSPDLGQTALSPDPSQES 1340

QY 814 ----- 814
DB 1341 SLSPDLGQTSLSPLDQESLSPDLGQTALSPDPSQESLSPDLGQTSLSPLDQESLSPDL 1400

QY 814 ----- 814
DB 1401 GQTALSPDLSQESLSPDLGQTALSPDLGQTALSPDLGQTALSPDLGQTALSPDLGQTALSPDL 1460

QY 814 ----- 814
DB 1461 GQLPLPERGQTPNADIGQMPSPDPSLNTMTFIDEEFNPLVYVGLSRDGDYIELIPRO 1520

QY 814 -----RSFQKTRHYFAAVER 830
DB 1521 KEESSEEDYGEFEYAYNDPIOTDLRTDINSRNDNTAAMTLKRSNTGKRYTTAAEET 1580

QY 831 LMDYKSSSPHYLRNRAAGS--VPQ--FKKVVFOETDGSFTQPLYRGLNELHGLL 884
DB 1581 SMDYS-----KFVQSDVDVYVEDIVYKVVYFRKYLDTFTRKLDPOGQEEHLGLL 1631

QY 885 GPYIAEVEDNIMVTFRNQAAPRYSFYSLSIE-----EDORQGAEPKKNFVKNPNE 936
DB 1632 GYVIRAEVDVYQVFEKNLASRYSILHAGLSYKESSEKTYEDDSPEWFKEDNAIOPNK 1691

QY 937 TKTYFMKVQHNAAPTKDEDDCKAMAYFSDVDLEKRVHSGLIGPLVCHTNTLNPAAHQV 996
DB 1692 TTYVWHATTNSGPENPGSACRAMAYYSANPEKDIHSGLIGPLLCRKGTLDKETNMPV 1751

PROSITE: PS00080; MULTICOPPER_OXIDASE2; 1.
DR PFM: PF00394; Cu-oxidase; 3.
KW OXIDOREDUCTASE; COPPER; METAL-BINDING; GLYCOPROTEIN; PLASMA; REPEAT;
KW SIGNAL; FOLYMOPHISM; 3D-STRUCTURE.
FT CHAIN 1
FT DOMAIN 20 1065 CERULOPLASMIN.
FT DOMAIN 20 357 F5/8 TYPE A 1.
FT DOMAIN 20 200 PLASTOCYANIN-LIKE 1.
FT DOMAIN 209 357 PLASTOCYANIN-LIKE 2.
FT DOMAIN 370 718 F5/8 TYPE A 2.
FT DOMAIN 370 560 PLASTOCYANIN-LIKE 3.
FT DOMAIN 570 718 PLASTOCYANIN-LIKE 4.
FT DOMAIN 730 1061 F5/8 TYPE A 3.
FT DOMAIN 730 900 PLASTOCYANIN-LIKE 5.
FT DOMAIN 908 1061 PLASTOCYANIN-LIKE 6.
FT CARBOHYD 138 138
FT CARBOHYD 358 358
FT CARBOHYD 397 397
FT CARBOHYD 762 762
FT DISULFID 174 200
FT DISULFID 276 357 PROBABLE.
FT DISULFID 534 560 PROBABLE.
FT DISULFID 637 718 PROBABLE.
FT METAL 120 120 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 122 122 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 180 180 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 182 182 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 994 994 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 997 997 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 999 999 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 1039 1039 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 1040 1040 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 1041 1041 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 1045 1045 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 1050 1050 COPPER (TYPE 1) (BY SIMILARITY).
FT VARIANT 79 79 T -> G.
FT VARIANT 449 449 L -> G.
FT VARIANT 1060 1060 E -> ESEYP (IN REF. 3).
FT CONFLICT 1065 1065
SQ SEQUENCE 1065 AA; 122205 MW; D127894A CRC32;
Query Match 22.38; Score 1741; DB 1; Length 1065;
Best Local Similarity 32.68; Pred. No. 8.5e-105;
Matches 389; Conservative 196; Mismatches 416; Indels 194; Gaps 23;
Db 5 LSTCFELCLRFCSATRRYYLGAVELSMDYQSDLGE---LPVDARPPRPVPSKFPENT 61
6 LGIFLFLCSTP-AWAKKEHYIGIETWDY-ASDHGEKLLISVDTEHSNITLONGPRI 63
Db 62 SVYTKTLFEFTDLENIARPRPMGLPPTIOAEYDTFVITLKNMASHPVSLAVG 121
64 GRLYKALYLDYDETFETTEKRPWLGLPILIKAEFGDKYVYHLKMLASRPYFHSHG 123
QY 122 VSYKASGAYDQOTSQREKEDKVFPGSGHTYVWYVLKENGPRASPLCLITSYLSHV 181
Db 124 ITYKHEGALYPNNTDORADKVPYEQYTYMLATEQSPGEGGNCVTRLYHSI 183
Db 162 DLVNDLNGSLGALLVREGSLAKEKQTL-HKFIPLAVFDEGSKSHSE-----TNS 234
Db 184 DAPDINAGLGLPLICKSDKEKEKHIDREFVYMSVYDENSWSLLEONIKYCSGP 243
QY 235 LMODRAASARAMPKMTVNGVYNSLPLGLIGCHRRSYVMVYVIGTTPPEVHSIFLEGHT 294
Db 244 EKVKNDNDEQESNMYSVNGYTFESLPLGSLACADRYKWLFGNGENVYHAAFEHGOA 303
QY 295 FLVRHHRASLEISPTFLAQTLMDGQFLPCHISSHODGHEAVYKDSCEEPQL 354
Db 304 LTNKVRIDTINLPATLPDIYMAQNGEMLSCONTLNHLKAGDAEFYQEC----- 358
QY 355 RKNNEAEADYDDDLTDEMDVRFDDDNPSFIIQIRSVAKKHPTWYHYIAAEEDMDY 414
Db 358 ---NKSSSKD-----NIRCKHVRH-----YIIAAEELIMNY 385

QY 415 APL-----VLAPDRSKSOYLNNNGPQIRGKRYKRYAFAYTDEF---KTREAIQ 462
Db 386 APSIDIFTEKNTLAPSDS--AVFEQGTTRIGSGKLYREYTPASFTNREKREE 443
QY 463 HESGILPILYGEADLLIFRKNQASRPYNYPHGI-----TDVRPLY---SRRLP 511
Db 444 EHLGILPILYAEVGDITIRYFNHKGAYPLSTFPIGVAFKNNNGITYSPRYNQSSVP 503
QY 512 KGVHLDFPLIPGEIFKRYKVTVEDEPTKSDRCLTRYSSFVNMERDLASGLIGPL 571
Db 504 PSASH-----VAPETFEYEMTVKVEGPTADPVCIAKMYSAVDPTKQIFGLIGPMK 558
QY 572 ICYEESDQKQNSDKRNVILFSVDENRSWTLTENIQFLEPNPAGVQLEDEPEQASN 631
Db 559 ICKKGSLEHANGROKDYKEFLPFTPEDEMSLLEEDINIRFTAPQVNDKEDDFQESN 618
QY 632 IMHSINGVFDLSQ-LASVCLHEVAWYIISGAQDPLVSFSGYTRFKKKVYEDTLTF 690
Db 619 KHSNMGFMYNQGLTCKGDSVWYILFSAGNADVHGITYSGNTYLMRGERDITANLF 678
QY 691 PFSGETYFMSNENPGLMILGCHNSDFRNRGTALLKYSKCDKNTGDIYEDYEDISAYLL 750
Db 679 PQTSLTLHMDTEGTFNVECLTDTHTGKKQKRYTNQCRQS-----EDS----- 726
QY 751 SKNNAIEPRSPQNPVYKLRHREITRTTLOSQDEEIDYDITISVEKKEDFDIYEDEN 810
Db 726 -----
QY 811 QSPRFQKTHRYFLAVERLMDYGMSSP-----HYLRNRAQSGV-----P 853
Db 726 ---TFIGERTYIYIAVEEMDY---SPQREKELHILQEDVNSAFLDKKEPIYGS 777
QY 854 QFRKVVQOEFTDGSFTQPLRYGELNEHLGLPYRAVEEDINIVTFRNQASRPYSFYS 913
Db 778 KKKVYVRYQYDSFRVYVEKAEENHGLIGPLDADVGQKVIIFKNMARTYSIHA- 837
QY 914 LISTEEDROGAEPRKRFVK---PNETKTYWKQHHMARTKDEDFCKAAVYSDVDLEK 970
Db 837 -----HGVQTESSVYPTLPGETLYVWKIPERSGAGTESACIPMAVYESTVDQVK 887
QY 971 DVHSLGLPLVCHTNTLNPAGROVYVOEALFFETFDKSKSYFMENMRNCRAPONI 1030
Db 888 DLVSLGLPLVCRPLIKYFNPRKL--EPALFLVFDENESYLDONITFYSDHPKV 945
QY 1031 QMEDPTEKNTFRFAINGYIDTLPGLVMAODORIRYTWLSGNSNENSHIFSGHVTY 1090
Db 946 NKDEEFIESKMAINRMGNGNLGILMHGDEVNMYLMGMGMEIDLHTVHFGHSFOY 1005
QY 1091 RKKEEYKALYNYPGVFETVEMLPKAGIWRVDECLGEHLHAGMSTFLVYSNK 1145
Db 1006 KHRGVYSSVDYDIPPGYQGLEMPRPRTGIWLILCHVTYDHIHAGMETTYVLONE 1060
RESULT 7
CERU_RAT STANDARD; PRT; 1059 AA.
AC P13635; 064719;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CERULOPLASMIN PRECURSOR (EC 1.16.3.1) (FERROXIDASE).
GN CP.
OS RATTUS NORVEGICUS (RAT).
OC EUCARYOTA; METAOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCURIONATHI; MORIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-LUNG, AND LIVER;
RX MEDLINE; 90237081.
RA FLEMING R.E., GITLIN J.D.;
RT "Primary structure of rat ceruloplasmin and analysis of
tissue-specific gene expression during development."

RL J. BIOL. CHEM. 265:7701-7707(1990).

RP [2] SEQUENCE OF 257-294; 571-612 AND 823-892 FROM N.A.

RC TISSUE=LIVER;

RA MEDLINE: 87137545.

RA ALFRED A.R., GRIMES A., SCHREIBER G., MERCER J.F.B.;

RT "Rat ceruloplasmin. Molecular cloning and gene expression in liver, choroid plexus, yolk sac, placenta, and testis."

RL J. BIOL. CHEM. 262:2875-2878(1987).

CC -1- FUNCTION: CERULOPLASMIN IS A BLUE, COPPER-BINDING (6-7 ATOMS PER MOLECULE) GLYCOPROTEIN FOUND IN PLASMA. FOUR POSSIBLE FUNCTIONS ARE FERROXIDASE ACTIVITY, AMINE OXIDASE ACTIVITY, COPPER TRANSPORT AND HOMEOSTASIS, AND SUPEROXIDE DISMUTASE ACTIVITY.

CC -1- FUNCTION: MAY ALSO PLAY A ROLE IN FETAL LUNG DEVELOPMENT OR PULMONARY ANTIOXIDANT DEFENSE.

CC -1- CATALYTIC ACTIVITY: 4 FE(2+) + 4 H(+) + O(2) = 4 FE(3+) + 2 H(2)O.

CC -1- COFACTOR: BINDS 6 CU-IONS PER MOLECULE.

CC -1- COFACTOR: THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE 3 OR COUPLED BINUCLEAR.

CC -1- TISSUE SPECIFICITY: SYNTHESIZED IN LIVER AND SECRETED INTO THE PLASMA. ALSO CHOROID PLEXUS, YOLK SAC, PLACENTA, AND TESTIS; NOT IN STOMACH AND SMALL INTESTINE. FETAL LUNG AND LIVER.

CC -1- INDUCTION: BY INFLAMMATION.

CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF 2 PLASTOCYANIN-LIKE REPEATS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: L33869; G499669; -

DR EMBL: M80529; G203421; -

DR EMBL: J02670; G203416; ALT_SEQ.

DR EMBL: M14102; G554429; -

DR PIR: A35210; A35210.

DR PIR: A29564; A29564.

DR PROSITE: PS00079; MULTICOPPER_OXIDASE1. 3.

DR PROSITE: PS00080; MULTICOPPER_OXIDASE2. 1.

DR PFAM: PF00394; Cu-oxidase; 3.

DR HSSP: P00450; 1KCW.

DR XINDREDCINASE; COPPER; METAL-BINDING; GLYCOPROTEIN; PLASMA; REPEAT; SIGNAL.

FT CHAIN 1 19 PROBABLE.

FT DOMAIN 20 1059 CERULOPLASMIN.

FT DOMAIN 20 356 F5/8 TYPE A 1.

FT DOMAIN 20 199 PLASTOCYANIN-LIKE 1.

FT DOMAIN 208 354 PLASTOCYANIN-LIKE 2.

FT DOMAIN 369 712 F5/8 TYPE A 2.

FT DOMAIN 369 554 PLASTOCYANIN-LIKE 3.

FT DOMAIN 564 710 PLASTOCYANIN-LIKE 4.

FT DOMAIN 724 1055 F5/8 TYPE A 3.

FT DOMAIN 724 894 PLASTOCYANIN-LIKE 5.

FT DOMAIN 902 1051 PLASTOCYANIN-LIKE 6.

FT DISULFID 173 199 BY SIMILARITY.

FT DISULFID 275 356 BY SIMILARITY.

FT DISULFID 528 554 BY SIMILARITY.

FT DISULFID 631 712 BY SIMILARITY.

FT DISULFID 868 994 BY SIMILARITY.

FT METAL 120 120 COPPER (TYPE 2) (BY SIMILARITY).

FT METAL 122 122 COPPER (TYPE 3) (BY SIMILARITY).

FT METAL 122 179 COPPER (TYPE 3) (BY SIMILARITY).

FT METAL 181 181 COPPER (TYPE 3) (BY SIMILARITY).

FT METAL 181 988 COPPER (TYPE 1) (BY SIMILARITY).

FT METAL 988 991 COPPER (TYPE 2) (BY SIMILARITY).

FT METAL 991 991 COPPER (TYPE 2) (BY SIMILARITY).

FT METAL 993 993 COPPER (TYPE 3) (BY SIMILARITY).

FT METAL 1033 1033 COPPER (TYPE 3) (BY SIMILARITY).

FT METAL 1034 1034 COPPER (TYPE 1) (BY SIMILARITY).

FT METAL 1035 1035 COPPER (TYPE 3) (BY SIMILARITY).

FT METAL 1039 1039 COPPER (TYPE 1) (BY SIMILARITY).

FT METAL 1044 1044 COPPER (TYPE 1) (BY SIMILARITY).

FT CARBOHYD 138 138 POTENTIAL.

FT CARBOHYD 226 226 POTENTIAL.

FT CARBOHYD 396 396 POTENTIAL.

FT CARBOHYD 582 582 POTENTIAL.

FT CARBOHYD 756 756 POTENTIAL.

FT CARBOHYD 920 920 POTENTIAL.

FT CARBOHYD 920 920 POTENTIAL.

FT CONFLICT 271 271 G -> A (IN REF. 2).

FT CONFLICT 604 605 ED -> DN (IN REF. 2).

FT CONFLICT 823 823 T -> S (IN REF. 2).

FT CONFLICT 833 833 V -> L (IN REF. 2).

FT CONFLICT 868 868 C -> V (IN REF. 2).

FT CONFLICT 891 891 L -> R (IN REF. 2).

SQ SEQUENCE 1059 AA; 120840 MW; 8E9F8FAD CRC32;

Query Match 21.7%; Score 1694; DB 1; Length 1059;

Best Local Similarity 32.5%; Pred. No. 9.2e-102;

Matches 386; Conservative 188; Mismatches 428; Indels 186; Gaps 23;

5 LSTCFELLCRCFSATRRYLGAVELSDYMO-SDLGEL-PVDARPPRPKSPFNTS 62

6 LSALEFL-HSLAMTRREKHYTGTEAWVDYASGSEKEELISVDTEQSNFYLRNPDHG 64

63 VYKRTLEVEFTDHLFNIAKPPRPMGLGPTIOAEVYDTVYITLKNASHPVSLHAGV 122

65 RYKRALYSEYDGTFTFKIDKPAWLGEPYKAKEVDKSHVHKNASRREYFHAGV 124

123 SYWKSSEGEYDDOTQOREKEDKVPGGSHYVQVLKENGPMASDPLCLYSLSHVD 182

125 TYTKAMEGALYDNTIDPRADKLPFGQOYLIVLRA-NEPSPGSDSCVRIYTHSHVD 183

183 LVKDLNSGLIGALLVCREGLAKERTQTL-HKFIILFAVFDGKSHSETKNSLM----- 237

184 AKPDIASGLIGLILCKKSLHKEKENIDQEFVLMFVSVDENLSWYLEDNITKTCSEPE 243

237 -QDRDASARANKPKHTVNGYVNRSLPGLIGCHRSYVHVYGMGTTPVYHSIFLEGHTF 295

244 KYDKNEDQESNRMYSSINGYTFGLPGLSMCAEDRVKVKYLLGKMGENVYHSELFHQAL 303

296 LVNRHQALESITPLEAQTLLMDLGOFLFCHSHSHOHGMEAYVYVDCSPBPQUR 355

304 TSKNYTTDITNLPATLLIVSMAQNPQWMLSCONLNLKGLQAFYVYRRC----- 357

356 MKNNEAEEDYDDLTLDSEMDVYRFDDNSPFIQIRSAKKHPRKTVWHYIAAEEDMDYA 415

357 ---NKPS- -DDDIQDRHV-----RH-----YIAAEETIMDYA 385

416 P-----LVLPDRSYKSOYLNNQPORIGKKYKVFMAATDFTF---KTRAI 461

386 PSGTDFTFGENTSLGSDSRVFEQ---GATRIIGSYKLVLYREYTDSTFNRRKERGD 441

462 QHESGLPLXGEVGDPTLLIFKNQASRPYNYPIGHTDVR---PLYSRLPKGVKHL 517

442 EEHLGLIGLYMAEVDIIRYTFHNKGOPPLSTOPGAVFTKENESTYGG---PDORSSK 498

518 KDPILPGLIFKKTVTVEDEGTPKCLTRYSSFYNNMERDLASGLIGPLITCKKS 577

499 QASHVAPKREFTYEMVVPKEMGPTADPVLCSKMYSGVDLTKDIFTGLIGPKKICKKS 558

578 VDRQGNQIMSDRKNNVLFVSFDENRSMWYLTENORFLPDPAGVQLDPEQASNNHGIN 637

559 LADGQRQKVDKRFYLFATVFDENESLLDDNTRMFTTAPENVYKDEDPQESKKNHSHN 618

638 GYVFDLSHQ-LSVLHVAWYIILSIGQTDPLSVFSGYTFKHAKVYEDTLFLPESGET 696

619 GPYAGNLPGLMGLGSIWYLFSAQNEADVHGIIYSGNITLSKGRDPTANLFPKSL 678

697 VFMSPENGLMILGCHNSDFRNKGATLKVSSCDKNTGDIYEDSTEDISAYILSKNNAL 756

679 LLMPTDEGSFDEVLCTTDHYTGMMOKQYTVNOC-----KGQFEDVLY----- 723

QY 757 EPRFSQNPVLKRRHQEIRITLQSDQDEIDDDITSEMKKEDPIYEDENQSPRSF 816
Db 723 ----- 723
QY 817 OKKTRHYFLAVERLMDYGMSSPHVLNRNAGSGV-----POFKRVYQOE 863
Db 723 -OGERHYTYAAVEVMDYSPSRDWMEMHLHLOQONSNAPLDEKFFPISGKRYKRYREF 781
QY 864 TDGSEFOPPLRGELNEHLGLGYIRAEVDNMTVFRNQAASPYSSLSIYEEDQ 923
Db 782 TDSTFEQVYKRRAREEHLGLIHLHADVAKVYKKNATRPYSIHA-----H 831
QY 924 GAPPKRNFKV---PNEKTYFMKVQHHMARTKDFPCAKAYFSDDVLEADVHGLGPL 980
Db 832 GVTKSTVAPPLPGEFRITYWQIPERSGAGTSDSPICPAPYXSTVDVADLYSGLGPL 891
QY 981 LVC---HTNLPANRQOVTVQEFALFEFTFEDTKSWYFENNERNRACNQMEDPTF 1037
Db 892 IYCRKSYVAVENK-----KKMEFSLFLVFDENESYLDNDINTYDPDEKDKKDEEF 946
QY 1038 KENYRFAINGYIMDTLPGLVMAQDRIKRYLLSMGSNENHSHFSGHYTVRKKKEEK 1097
Db 947 IESNKKHAIKMGKMGGLQGLTMHGVDEVNMYVMANGNEIDLHTVHFHGSFOYKRGHIS 1006
QY 1098 MALYNLYPGVFETVEMLPKAGIMRVBCLIGELHAGMSTLFVYSK 1145
Db 1007 SDVFDEFPQYQTLMEFPOTPGTWLHCHVTDHIGHAMVTVTVLPNQ 1054
RESULT 8
CERU MOUSE STANDARD: PRT: 1062 AA.
ID CERU MOUSE
AC 061147;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CERULOPLASMIN PRECURSOR (EC 1.16.3.1) (FERROXIDASE).
GN CP.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MORINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RL KLOMP L.W.J., FARHANGRAZI Z.S., CHOI D.W., GITLIN J.D.;
RN SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RX TISSUE SPECIFICITY.
RA MEDLINE: 96394736.
KLOMP L.W.J., FARHANGRAZI Z.S., DUGAN L.L., GITLIN J.D.;
Ceruloplasmin gene expression in the murine central nervous system.";
J. CLIN. INVEST. 98:207-215(1996).
-1- FUNCTION: CERULOPLASMIN IS A BLUE, COPPER-BINDING (6-7 ATOMS PER
MOLECULE) GLYCOPROTEIN FOUND IN PLASMA. FOUR POSSIBLE FUNCTIONS
ARE FERROXIDASE ACTIVITY, AMINE OXIDASE ACTIVITY, COPPER TRANSPORT
AND HOMEOSTASIS, AND SUPEROXIDE DISMUTASE ACTIVITY.
-1- CATALYTIC ACTIVITY: 4 FE(2+) + 4 H(+) + O(2) -> 4 FE(3+) + 2 H(2)O.
-1- COFACTOR: THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH
CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE
2 OR NORMAL, AND TYPE 3 OR COUPLED BINUCLEAR.
-1- TISSUE SPECIFICITY: MANY TISSUES, INCLUDING LIVER, EYE AND BRAIN.
-1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
2 PLASTOCYANIN-LIKE REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL: U49430; G1224108; -

DR MGD: MG1:88476; CP.
DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE: PS00080; MULTICOPPER_OXIDASE2; 1.
DR PFAM: PF00394; Cu-oxidase; 3.
DR HSSP: P00450; 1KCW.
KW OXIDOREDUCTASE; COPPER; METAL-BINDING; GLYCOPROTEIN; PLASMA; REPEAT;
KW SIGNAL.
FT SIGNAL. 1 19
FT CHAIN 20 1062
FT DOMAIN 20 199
FT DOMAIN 208 356
FT DOMAIN 369 713
FT DOMAIN 565 725
FT DOMAIN 725 1057
FT DOMAIN 1057 1086
FT DOMAIN 1086 1097
FT DISULFID 173 199
FT DISULFID 275 356
FT DISULFID 529 555
FT DISULFID 632 713
FT DISULFID 870 896
FT METAL 120 120
FT METAL 122 122
FT METAL 179 179
FT METAL 181 181
FT METAL 990 990
FT METAL 993 993
FT METAL 1035 1035
FT METAL 1036 1036
FT METAL 1041 1041
FT METAL 1046 1046
FT CARBOHYD 138 138
FT CARBOHYD 226 226
FT CARBOHYD 383 383
FT CARBOHYD 625 625
FT CARBOHYD 757 757
FT CARBOHYD 922 922
SO SEQUENCE 1062 AA; 121159 MW; 94693E7 CRC32;
Query Match 21.2%; Score 1651; DB 1; Length 1062;
Best local Similarity 32.3%; Pred. No. 5,5e-99;
Matches 382; Conservative 196; Mismatches 443; Indels 160; Gaps 25;
QY 10 FLICLRFCF-----SARRRYLAGAVELSDWYMO--SDLGELPVDAKPPRYPKSFPFN 60
Db 3 FLLSTFIFLYSSIALARDKHYFIGITEAVMDYASGTEKKLLSVDEQSNFVLQNGPDR 62
QY 61 TSVYKKTLEFVETDHLFNIAKRPKRMGLGPTIAEYDVYVITLKNASHPVSLHAY 120
Db 63 IGRKRYKALFEYTDFTSKTIDKPAMLGFLGVIAEVDKYYVHLKNLASRYTFHAH 122
QY 121 GVSVMKASEGAEVDQTSQREKEDDKVFCGSHTYWQVLEKNGPMASDPLCTGYSLSH 180
Db 123 GVTYTYEBAVYFDPVTDTPQKADKVLVHQQYVYLHA-NESSPGEBSNCTRIYHSH 181
QY 181 VDLVKDNLGSLGALLVCREGSLAKERTQTL-HKFTLFAVFEDEGKSWHSETKNSLM-- 237
Db 182 VDAKDIASGLIPLILCKKGSLYKKEKKIIDQEFVLMFSDVDENLSWYLEDNIKTFCS 241
QY 237 ---QDDAASARAMPKMHVNGVYNSLPLJICHRKSVYWHYIGCTTPEVHSIFLEGH 293
Db 242 PEKVYDQNEDEFSNMYSTINGTFSGLSMCAADRYAKWYLFEGMGNEVYDVSAPFHQ 301
QY 294 TFLVRNHRQASLEISITFTLTAOTLMDLQOFLFLFCHISSHQHDEAVYKVCCEPEQ 353
Db 302 ALSRMYQYDIDILFPAITLIDAYVANOQGVWMLSCQNLNHLKAGIQAQAFQVQDCC----- 357
QY 354 LRRKNNDEADYDDDLTDSEMDVYRFDDDNSPSFIOIRSVAKRHKPTWYIAAEEDMD 413

316 KINMATAOSNAKEMELQVLDLCTQKVTGIIQAGANDFGHIQVASYKVAHSDGQVMTVY 375
QY 1400 FQNGKVFQGNODSFTPVVNSLPPLTRRLRHPOSWHOIALRMEVLCG 1451
DB 376 EEOGTSKVFQGNLDNNSHKNIFEKPFMARYRVLPPLSMNRITRLRLGLCG 427

RESULT 10
MEGM_MOUSE STANDARD: PRT; 463 AA.

AC P21956; P97800;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE MILK FAT GLOBULE-EGF FACTOR 8 PRECURSOR (MEG-E8) (MEGM) (SPERM SURFACE
PROTEIN SP47) (PP47).
GN MEG8.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIRURGONATHI; MORIDAE; MURINAE; MUS.
[1]
SEQUENCE FROM N.A. AND SEQUENCE OF 23-35.
TISSUE-MAMMARY GLAND;
RX MEDLINE; 91046008.
RA STUBBS J.D., LEKUTIS C., SINGER K.L., BUI A., YUZUKI D.,
SRINIVASAN U., PARRY G.;
RT "DNA cloning of a mouse mammary epithelial cell surface protein
reveals the existence of epidermal growth factor-like domains linked
to factor VIII-like sequences.";
RL PROC. NATL. ACADE. SCI. U.S.A. 87:8417-8421(1990).
[2]
RN SEQUENCE OF 23-456 FROM N.A.
RC TISSUE-TESTIS;
RA ENSLIN M.A.;
RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC - FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
BINDING PROTEIN.
CC - SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC - TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
SPERMATOZOAN.
CC - DEVELOPMENTAL STAGE: MRNA EXPRESSION IS DETECTABLE IN MAMMARY
TISSUE FROM NONPREGNANT ANIMALS & MAXIMAL IN THE LACTATING GLAND.
CC - SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC - SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M38337; G199143; -
CC EMBL; Y11684; E307035; -
CC PIR; A36479; A36479; MEG8.
CC MGD; MGI:102768; MEG8.
CC PROSITE; PS00022; EGF_1; 2.
CC PROSITE; PS01285; EGF_2; 2.
CC PROSITE; PS01285; FA58C_1; 2.
CC PROSITE; PS01286; FA58C_2; 2.
CC PRAM; PR00008; EGF_2.
CC PRAM; PR00754; F5_F8_Type_C; 2.
CC HSP; P00740; 11X.
CC SIGNAL; GLYCOPROTEIN; REPEAT; EGF-LIKE DOMAIN; MILK.
CC KW SIGNAL.
CC FT CHAIN 1 22
CC FT 1 22
CC FT 23 463 MILK FAT GLOBULE-EGF FACTOR 8.
CC FT 24 61 EGF-LIKE 1.
CC FT 64 108 EGF-LIKE 2.
CC FT 148 303 F5/8 TYPE C 1.
CC FT 308 463 F5/8 TYPE C 2.
CC FT 87 89 CELL ATTACHMENT SITE (POTENTIAL).
CC FT SITE 87 89
CC FT DISULFID 28 39 BY SIMILARITY.

FT DISULFID 33 49 BY SIMILARITY.
FT DISULFID 51 60 BY SIMILARITY.
FT DISULFID 68 79 BY SIMILARITY.
FT DISULFID 73 96 BY SIMILARITY.
FT DISULFID 98 107 BY SIMILARITY.
FT DISULFID 148 303 BY SIMILARITY.
FT DISULFID 290 294 BY SIMILARITY.
FT DISULFID 308 463 BY SIMILARITY.
FT CARBOHYD 61 61 POTENTIAL.
FT CARBOHYD 266 266 POTENTIAL.
FT CARBOHYD 316 316 POTENTIAL.
FT CARBOHYD 426 426 POTENTIAL.
FT CONFLICT 30 30 S -> F (IN REF. 2).
FT CONFLICT 35 35 N -> D (IN AA SEQUENCE; REF. 1).
FT CONFLICT 55 55 F -> S (IN REF. 2).
FT CONFLICT 67 67 P -> G (IN REF. 2).
FT CONFLICT 110 147 ETNYNDGEYMTFAVNPNTAVPTAPTPDLSNNLSR ->
G (IN REF. 2).
FT CONFLICT 168 168 Y -> S (IN REF. 2).
FT CONFLICT 196 196 H -> T (IN REF. 2).
FT CONFLICT 309 309 L -> S (IN REF. 2).
FT CONFLICT 394 395 VE -> AQ (IN REF. 2).
FT CONFLICT 410 410 V -> D (IN REF. 2).
SQ SEQUENCE 463 AA; 51465 MM; 254BAB08 CRC32.

Query Match 8.4%; Score 657; DB 1; Length 463;
Best Local Similarity 42.2%; Pred. No. 2,9e-35;
Matches 135; Conservative 63; Mismatches 110; Indels 12; Gaps 5;

QY 1143 SNKQTPUGMASGIRPQTTASQY-----GOMAKRLAHYSGSINM--STKEPFS 1194
DB 145 ASRCSTQGMGGALADSDASVYVGMFGALQRMGPDLRLRYTGIVNMAHNSYDSKP 204
QY 1195 WIKDLAPMIHIGIKQROKSFSLYISOFITMSLDKMKOTYRGNSGTLMVFFGN 1254
DB 205 WIOVNLKRVKSVQMTQGSRRGRACYLKTIFYANSLDKRKEFIDESGDD-KEFLGN 263
QY 1255 VDSGGINHNFPIIARYRLPHPTYSIRSTLMEIMGCDLNSCMPLGMSKAISDAQ 1314
DB 264 LDNNSLKVNENPFLQAYIRLVPVCHRCITLRFELGELMGCLPELDKNNITPDSQ 323
QY 1315 ITASSTY--NMA-TWSPKARLHLOGNSAMPPOVNNPKFQLQVDFQTKMYGVYTTQ 1371
DB 324 MSASSSYKTNLRAFGWYPHGLGRDNGKINAWTAQNSAKEMELQVLDLGTQRYTGITQ 383
QY 1372 GVSLSLTSMVYKELISSDQDGHMTLFFQNGKYVQGNDSFTPVVNSLPPLTRRL 1431
DB 384 GARDFGHIQVSEYKVAHSDGQVMTVEEOGSSKVFQGNLDNNSHKNIFEKPFMARY 443
QY 1432 RIHPQSWHOIALRMEVLCG 1451
DB 444 RVLPSVSMNRITRLRLGLCG 463

RESULT 11
MEGM_PIG STANDARD: PRT; 409 AA.

AC P79385;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE MILK FAT GLOBULE-EGF FACTOR 8 (MEG-E8) (MEGM) (SPERM SURFACE PROTEIN
SP47) (PP47).
GN MEG8.
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
[1]
SEQUENCE FROM N.A.
TISSUE-TESTIS;
RA ENSLIN M.A.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

Query Match 8.18; Score 635; DB 1; Length 427;
Best Local Similarity 37.98; Pred. No. 6.8e-34;
Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7,

	RESULT	13
MEGM_HUMAN	MEGM_HUMAN	
ID	MEGM_HUMAN	STANDARD:
AC	008431:	PRT: 387 AA.
DT	01-OCT-1996 (REL. 34, CREATED)	
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)	
DT	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)	
DE	MILK FAT GLOBULE-EGF FACTOR 8 PRECURSOR (MEG-E8) (HMFG) (BREAST EPITHELIAL ANTIGEN BA46) (MEGM).	
DE	MEG8.	
GN	HOMO SAPIENS (HUMAN).	
OS	HOMO SAPIENS (HUMAN).	
OC	EUMARIOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;	
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.	
RN	[1]	
RN	SEQUENCE FROM N.A.	
RC	TISSUE-BREAST, AND BREAST CARCINOMA;	
RC	MEDLINE; 96213908.	
RA	CUTOO J.R., TAYLOR M.R., GODWIN S.G., CERIANI R.L., PETERSON J.A.;	
RT	"Cloning and sequence analysis of human breast epithelial antigen BA46 reveals an RGD cell adhesion sequence presented on an epidermal	

QY	1284	RSTLMEMLGCDLNSCSMPLGMSKAISDAQITASSYF	7.5%; Score 588; DB 1; Length 387; Best Local Similarity 37.3%; Pred. No. 6.5e-31; Matches 132; Conservative 69; Mismatches 125; Indels 28; Gaps 9;	growth factor-like domain"; DNA CELL BIOL. 15:281-286(1996). [2] SEQUENCE OF 170-387 FROM N.A. RP TISSUE-MAMMARY GLAND; RC MEDLINE: 91371351. RA LAROCCA D., PETERSON J.A., URREA R., KUNIOYOSHI J., BISTRAN A.M., CERIANI R.L.; "A rat 46,000 human milk fat globule protein that is highly expressed in human breast tumors contains factor VIII-like domains."; RT CANCER RES. 51:4994-4998(1991). RL -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN. CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES. CC OVEREXPRESSED IN SEVERAL CARCINOMAS. CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN. CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS. ----- CC This SWISS-PROT entry is copyright. It is produced through a collaboration CC between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See http://www.isb-sib.ch/announce/ CC or send an email to license@isb-sib.ch). ----- CC EMBL: U58516; GI381162; - DR EMBL: S56151; G235397; -. DR MIM: 602281; -. DR PROSITE: PS00022; EGF_1; 1. DR PROSITE: PS01186; EGF_2; 1. DR PROSITE: PS01285; FA58C_1; 2. DR PROSITE: PS01286; FA58C_2; 2. DR PFAM: PF00008; EGF; 1. DR PFAM: PF00754; F5_F8_type_C; 2. KM SIGNAL: GLYCOPROTEIN, MILK, REPEAT, EGF-LIKE DOMAIN. FT SIGNAL 1 22 POTENTIAL. FT CHAIN 23 387 MILK FAT GLOBULE-EGF FACTOR 8. FT DOMAIN 70 225 EGF-LIKE. FT DOMAIN 230 387 F5/8 TYPE C 1. FT DOMAIN 46 48 F5/8 TYPE C 2. FT SITE 46 48 CELL ATTACHMENT SITE (POTENTIAL). FT DISULFID 27 38 BY SIMILARITY. FT DISULFID 32 55 BY SIMILARITY. FT DISULFID 57 66 BY SIMILARITY. FT DISULFID 70 225 BY SIMILARITY. FT DISULFID 212 216 BY SIMILARITY. FT DISULFID 230 387 BY SIMILARITY. FT CARBOHYD 238 238 POTENTIAL. FT CARBOHYD 325 325 POTENTIAL. FT CARBOHYD 329 329 POTENTIAL. FT CARBOHYD 350 350 POTENTIAL. SQ SEQUENCE 387 AA; 43123 MW; 9672347E CRC32;
QY	1112	EMLPKAGIWRVECLIGEHLHAGNSTLFLVYSNKCQTPLGMASGHIRDPQTASG	1167	
DB	48	DVFPSP-----YTCCTCLKG--YAGNHC-----ETKCEVPELGMENGINANSQIAASSVRYTF	95	
QY	1167	-QYGMARLALRHSGSINM--STKEPFSIKIKDPLAPMIIHRIKQNGAQRKSSLI	1223	
DB	96	IGLQHWPELALNLNAGVANNATPSSNDNPNIYVNLRRMVGTVGVAQSRSLASHEYL	155	
QY	1224	SOFIMYSLDGKKMOTYKGNSTGTLMVFEFGVANDSSGICKHNIENPPIIARIYILPPTHYSI	1283	
DB	156	KAFKAIYASINGHEFD-FIHDAVKKRKEFGVGNKKNAVHNLEFETVEAOYVRLYFTSCHI	214	
QY	1284	RSTLMEMLGCDLNSCSMPLGMSKAISDAQITASSYF-----TNMFATWSPSKARLHLQG	1339	

DB 215 ACTURELLGCELNGCANPLGLKNNIPDKOITASSSYKTWGLHF-SWNSPYARLDKOG 273
QY 1340 RSNAMPQVNNPKEMVLOVDQTKMTKVTGVTGKVSLLTMYKEFLISSODGHWITLF 1399
FT DISULFID 145 171 PROBABLE.
FT DISULFID 204 226 PROBABLE.
FT DISULFID 273 422 BY SIMILARITY.
FT DISULFID 429 581 BY SIMILARITY.
SQ SEQUENCE 914 AA; 102480 MW; 59387139 CRC32;
QY 1400 F--ONGKRVKVGQNSDFTFPVNSLDPPLLTTRYLRHPOSVMQIALRMEVLGC 1451
DB 334 QDPRGTSSKIFPGNMWDSHKKNLFTPLIARYRLPVPAMHRIALRLLELGC 387
RESULT 14
NRP_CHICK STANDARD; PRT; 914 AA.
AC P79795;
DT 01-NOV-1997 (REL. 35, CREATED)
RT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
QY 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
NEUROFILIN PRECURSOR (A5 PROTEIN).
NRP.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
NC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEGHORN; TISSUE-EMBRYONIC BRAIN;
RX MEDLINE: 95324761.
RA TAKAGI S., KASUYA Y., SHIMIZU M., MATSURA T., TSUBOI M., KAWAKAMI A.,
FUJISAWA H.;
RT "Expression of a cell adhesion molecule, neuropilin, in the
developing chick nervous system."
RL DEV. BIOL. 170:207-222(1995).
CC -1- FUNCTION: CALCULUM-INDEPENDENT CELL ADHESION MOLECULE THAT FUNCTION
DURING THE FORMATION OF CERTAIN NEURONAL CIRCUITS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: DEVELOPING NERVOUS SYSTEM; OPTIC TECTUM
(LAYERS D AND E OF SEGS); AMACRINE CELLS OF RETINA, NEURITES OF
DORSAL ROOT GANGLIA. ALSO EXPRESSED IN NONNEURONAL CELLS, E.G.
CC BLOOD VESSELS IN THE ENTIRE EMBRYO.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D45416; G1841311; -
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01285; FA58C_1; 2.
DR PROSITE: PS01286; FA58C_2; 2.
DR PROSITE: PS00740; MAM_1; 1.
DR PROSITE: PS00740; MAM_2; 1.
DR PFAM: PF00431; CUB; 2.
DR PFAM: PF00629; MAM; 1.
DR PFAM: PF00754; F5_F8 TYPE C; 2.
KW TRANSMEMBRANE; GLYCOPROTEIN; NEURONE; SIGNAL; REPEAT; CELL ADHESION.
FT SIGNAL 1 18
FT CHAIN 19 914
FT DOMAIN 20 847
FT TRANSMEM 848 870
FT DOMAIN 871 914
FT DOMAIN 914 914
FT DOMAIN 145 263
FT DOMAIN 273 422
FT DOMAIN 429 581
FT DOMAIN 636 801
FT DISULFID 25 52
FT DISULFID 80 102
PROBABLE.
EMBL: D50086; E293356; -

FT DISULFID 145 171
FT DISULFID 204 226
FT DISULFID 273 422
FT DISULFID 429 581
SQ SEQUENCE 914 AA; 102480 MW; 59387139 CRC32;
Query Match 5.9%; Score 458.5; DB 1; Length 914;
Best Local Similarity 35.0%; Pred. No. 5.4e-22;
Matches 113; Conservative 58; Mismatches 129; Indels 23; Gaps 13;
QY 1145 KCOTPLGMASSGHRDFOQTASGOYGO-WAPKRLAHYSSSIANSTKEPF--SWIKYLL 1201
DB 272 QCEPMEGSEGHSDQITVSSQYSAINSSERSRLNYPE--NGWTPGEDSVREWIQVDIG 329
QY 1202 APMLIGITGGA--RQKSSLYISOFIMYSLDCKKQOTYR-GNSTGLMVFEGVDS 1258
DB 330 LRFVSGITGGAISKEIKYKTYRYDVSSNGEDWTLKEGKP--VYFGNSNPT 386
QY 1259 GIKHNIENPPIIARYIRLPHYSIRSLRMELGCDLNS--CSMPLGNSKAISDAQIT 1316
DB 387 DYYRFPKPVLTFRFVRIRKPVSEMGVSLRFVYCKITDPCSGMLGWSGLIDDSQIT 446
QY 1317 ASSYFTNMTATSPSKARHLQGRSNAMPQVNNP--KEMLVVDQTKMTKVTGVTGQVK 1374
DB 447 AS--TQVDRNMPENARL-ITRSQWALPPTHTPTNEMLDIDGEEKIVGIGIVQGG 502
QY 1375 SLTTSKYVEKEFLISSODGHWITLFEONG--RVKVGQNSDFTFPVNSLDPPLLTTRYLR 1432
DB 503 HRENKTFMKFKFGYNNNSDKMKIMDSKKIKTEGNTNDTDLPLRFE--PVSTRIR 561
QY 1433 IHPQSVHCO-IALRMEVLCEAO 1454
DB 562 VYPERATHAGLGRMELGCELE 584
RESULT 15
NRP_MOUSE STANDARD; PRT; 923 AA.
AC P97333;
DT 01-NOV-1997 (REL. 35, CREATED)
RT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE NEUROFILIN PRECURSOR (A5 PROTEIN).
GN NRP.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
NC RODENTIA; SCIROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-EMBRYONIC BRAIN;
RX MEDLINE: 96353149.
RA KAWAKAMI A., KITSUKAWA T., TAKAGI S., FUJISAWA H.;
RT "Developmentally regulated expression of a cell surface protein,
neuropilin, in the mouse nervous system."
RL J. NEUROBIOL. 29:1-17(1996).
CC -1- FUNCTION: CALCULUM-INDEPENDENT CELL ADHESION MOLECULE THAT FUNCTION
DURING THE FORMATION OF CERTAIN NEURONAL CIRCUITS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: NERVOUS SYSTEM.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D50086; E293356; -

DR MGD: MGI:106206; NRP.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FAS8C_1; 2.
DR PROSITE; PS01286; FAS8C_2; 2.
DR PROSITE; PS00740; MAM_1; 1.
DR PROSITE; PS50060; MAM_2; 1.
DR PFAM; PF00431; CUB; 2.
DR PFAM; PF00629; MAM; 1.
DR PFAM; PF00754; F5_F8_Type_C; 2.
DR TRANSMEMBRANE; GLYCOPROTEIN; NEURONE; SIGNAL; REPEAT; CELL ADHESION.
KM SIGNAL 1 20
FT CHAIN 21 923
FT DOMAIN 21 856
FT TRANSMEM 857 879
FT DOMAIN 880 923
FT DOMAIN 27 141
FT DOMAIN 147 265
FT DOMAIN 275 424
FT DOMAIN 431 583
FT DOMAIN 645 811
FT DISULFID 27 54
FT DISULFID 82 104
FT DISULFID 147 173
FT DISULFID 206 228
FT DISULFID 275 424
FT DISULFID 431 583
FT CARBOHYD 150 150
FT CARBOHYD 261 261
FT CARBOHYD 300 300
FT CARBOHYD 522 522
FT CARBOHYD 842 842
SQ SEQUENCE 923 AA; 103019 MW; 265CODE7 CRC32;

Query Match 5.7%; Score 446.5; DB 1; Length 923;
Best Local Similarity 34.7%; Pred. No. 3.2e-21;
Matches 112; Conservative 57; Mismatches 131; Indels 23; Gaps 13;
QY 1145 KCQTPPLMASGHTRDFOITASGQYG-QWAPKLARLHSGSINAWSTKEPF-SWIKVDL 1201
DB 274 KCEMALGMESEIHSDQITASQGTWVSVERSLNYPE-NGWTPGEDSYKEWIQVDLG 331
QY 1202 APMIIGHIKTOGA--RQKFSLSYISQFLIMYSLGKKKQOTYR-GNSTGTLVFFGNVDS 1258
DB 332 LRFVTAVGTGQAIKSKETKKKYVKTFRVDISSNGEDMISLKEGKA--LIIFQNTNPT 388
QY 1259 GIKHNIFNPPIIARYIRLHPHYISIRSTLRMLGCDLNS--CSMPLGMEKASDAQIT 1316
DB 389 DVLIGVFYSKPLITRFVRIKPYSWETGISMRVEYGCKITDIPCSGMLGNVSGLISDSQIT 448
DB 1317 ASSYFTNMFAWTSPSKARLHLGSRNAMPQVNNP--KEWLQVDFOKTMKVTGYTQGVK 1374
DB 449 ASNQADR--NMWPEINRL-VTSRTGVALPPSPHPYTNEMQLQVDLGDEKIVRGVILIOGK 504
QY 1375 SLTSMYVKEFLISSQDGHQWTLFPONG--KYKVFQGNDSFTPVVNSLDPLLTRILR 1432
DB 505 HREKRVFMREKFKIAYSNNGSDPKTIMDSKRKAKSFGNNNYDPPELRTFS-PLSTRIR 563
QY 1433 IHPOSWVHO-IALRMEVIGCEAQ 1454
DB 564 IYPERATHSGIGLRLMELLGCEVE 586

Search completed: August 19, 1999, 12:39:19
Job time: 2333 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 1999, 12:01:45 ; Search time 19.9 Seconds

(without alignments)

4505.984 Million cell updates/sec

Title: US-09-001-039A-47

Perfect score: 7
Score: 1

Sequence: 1 MQIELSTCFELCLRFCSA.....WHQIALRMEVLGCEAODLY 1457

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database

```

1: SPTRMBL.10.*
2:   sp_archaea.*
3:   sp_bacteria.*
4:   sp_fungi.*
5:   sp_human.*
6:   sp_invertebrate.*
7:   sp_mammal.*
8:   sp_mhc.*
9:   sp_organelle.*
10:  sp_plant.*
11:  sp_proct.*
12:  sp_virus.*
13:  sp_vertebrate.*
14:  sp_unclassified.*

```

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description
1	6347	81.4	2343	6	018806	018806	canis famlil	
2	6306	80.9	2343	6	0627330	0627330	canis famlil	
3	2387.5	30.6	2224	4	0473737	0473737	homo sapien	
4	2374.5	30.5	2183	11	08878783	08878783	mus musculul	
5	1808	23.2	1157	11	0987024	0987024	mus musculul	
6	1341	17.2	891	4	075180	075180	homo sapien	
7	1230	15.8	782	4	075659	075659	homo sapien	
8	1104	14.2	216	4	014286	014286	homo sapien	
9	662.5	8.5	480	4	043854	043854	homo sapien	
10	655.5	8.4	480	11	035474	035474	mus musculul	
11	514.5	6.6	333	6	077718	077718	equus caball	
12	469.5	6.0	931	4	060462	060462	homo sapien	
13	469.5	6.0	909	4	014820	014820	homo sapien	
14	469.5	6.0	926	4	014821	014821	homo sapien	
15	464.5	6.0	925	11	035276	035276	rattus norv	
16	462.5	5.9	909	11	035373	035373	mus musculul	
17	462.5	5.9	926	11	035374	035374	mus musculul	
18	462.5	5.9	931	11	035375	035375	mus musculul	
19	462.5	5.9	901	11	035376	035376	mus musculul	
20	462.5	5.9	906	11	035377	035377	mus musculul	
21	462.5	5.9	914	11	035378	035378	mus musculul	
22	429.5	5.5	923	4	060461	060461	homo sapien	
23	429.5	5.5	923	4	014786	014786	homo sapien	
24	408.5	5.2	558	5	076470	076470	lythechinus	
25	271	3.5	1128	11	088442	088442	mus musculul	
26	266	3.4	764	11	054860	054860	mus musculul	
27	262	3.4	85	4	014113	014113	homo sapien	
28	247	3.2	224	4	092614	092614	mus musculul	
29	243	3.1	224	4	015537	015537	homo sapien	

ALIGNMENTS

ID	Accession	Species	Length	Score	DB	Length	Score	DB
018806	PRELIMINARY	PRT: 2343 AA.						
018806	018806	018806	018806	018806	018806	018806	018806	018806
01-JAN-1998	(TREMblrel. 05, Created)							
01-JAN-1998	(TREMblrel. 05, Last sequence update)							
01-MAY-1999	(TREMblrel. 10, Last annotation update)							
DE	FACTOR VIII.							
GN	F8.							
OS	Canis familiaris (Dog)							
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;							
OC	Eutheria; Carnivora; Fissipedia; Canidae; Canis.							
RN	(1)							
RP	SEQUENCE FROM N.A.							
RC	TISSUE-LAYER:							
RA	CAMERON C., NOTLEY C., HOYLE S., MCGLYNN L., HOUGH C., KAMISUE S.,							
RA	GILES A., LILLICRAP D.,							
RL	Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.							
DR	EMBL; AF016234; AAB87412.1; .							
DR	PFAM; PF00394; Cu-oxidase; 3.							
DR	PFAM; PF00754; F5_F8_type_C; 2.							
DR	PROSITE; PS01285; FA58C.1; 2.							
DR	PROSITE; PS01286; FA58C.2; 2.							
SO	SEQUENCE 2343 AA; 265829 MW; 97COAF09 CRC32;							
Query Match	81.4%; Score 6347; DB 6; Length 2343;							
Best Local Similarity	53.6%; Pred. No. 0;							
Matches 1259; Conservative	83; Mismatches 105; Indels 900; Gaps							
QY	1	MOELSTCFELCLLRFCSATRRYYLGAVELSMDYMSD-IGELPYDARPPRPKSPFP	59					
DB	1	MOVELYTCFCFLDLPFSLSATRRYYLGAVELSMDYMSDLSALHADTSFSRVPGLPL	60					
QY	60	NTSVYVKKTLPEEFDDHFNIAKPRPPMGLGPIQAEVYTYVITLKNMASHVPSLHA	119					
DB	61	TTSVYVKKTLPEEFDDHFNIAKPRPPMGLGPIQAEVYTYVITLKNMASHVPSLHA	120					
QY	120	VGVSIVKMSSEGALEYDDQTSQREKEDDKYFPGSGSHYYVQVLKENGPMASDPLCTIYSYS	179					
DB	121	VGVSIVKMSSEGALEYDDQTSQREKEDDKYFPGSGSHYYVQVLKENGPMASDPLCTIYSYS	180					
QY	180	HVDLVKINDNSGIGALLVCREGSLAKERTQTLHKFILLFAVFEDEKSWHSETKNSLMODR	239					
DB	181	HVDLVKINDNSGIGALLVCREGSLAKERTQTLHKFILLFAVFEDEKSWHSETKNSLMODR	239					
QY	240	DAASRAMPKMHVNGVYNRSLPGIIGCHRSVYVHVIGMGTTPVHSTFLEGHTFLVYN	299					
DB	239	DAASRAMPKMHVNGVYNRSLPGIIGCHRSVYVHVIGMGTTPVHSTFLEGHTFLVYN	294					
QY	300	HROASLESPITFLTAQVLLMDLGOFLFCHTSSHHOGBMEAYVAVDSCPEEPQLRMKN	359					

Db 295 HRQASLEISPIFLTAQFLMADLGOFLFCHIPSQHODMEAYKVDCSPEEPOLRMKN 354
Qy 360 EEAEDYDDDLJDSEMDVYRFDDDNBSFQIOTSVAKKHKKTKTVHVIIAEEDMDVAPLV 419
Db 355 ED-KYDDDLJSDMDVYFDDDNBSFQIOTSVAKKHKKTKTVHVIIAEEDMDVAPLV 413
Qy 420 ADDDSYKSOYLNNGPORGKRYKKVRFMAVYDETFKTEALQIESGILGPLLYGEVDT 479
Db 414 TPNDSHKRLYLNNGPORGKRYKKVRFMAVYDETFKTEALQIESGILGPLLYGEVDT 473
Qy 480 LLIIFKNQASRPYNYTPHGITDVRPLYSRLPKGVKHLKDFPLPGELFKYKWTVEDEG 539
Db 474 LLIIFKNQASRPYNYTPHGINVTPLHGTGRLPKGVKHLKDMDFLPGELFKYKWTVEDEG 533
Qy 540 PKRSDPCLTRYSFVFNMERDLASGLIGPLLYCKESVDONGOMSKRVVILFESVD 599
Db 534 PKRSDPCLTRYSFVFNMERDLASGLIGPLLYCKESVDONGOMSKRVVILFESVD 593
Qy 600 ENRSWLTENIORFLPNPAVOLEDEPEFQASNIIMHSINGVDSIQLSYCLHEVAYWYL 659
Db 594 ENRSWLTENIORFLPNPAVOLEDEPEFQASNIIMHSINGVDSIQLSYCLHEVAYWYL 653
Qy 660 SUGAOTDELVSFSGYTFKHKMYEDTLPLPSPGETVMSKBNPGLWILGCHNSDFNRK 719
Db 654 SUGAOTDELVSFSGYTFKHKMYEDTLPLPSPGETVMSKBNPGLWILGCHNSDFNRK 713
Qy 720 GMTALLKVSCKNTGDYEDSYEDI:SAVLLSKNNAIEPRSF----- 762
Db 714 GMTALLKVSCKNRIDYEDYEDIEDI:PTPLNENNYIKRFSQNSRHSSTEKOLKAT 773
Qy 762 ----- 762
Db 774 TPENDIEKIDQSGERTQOLIKAQSVSSDLMLLQONPTPGLFLSLEADTBRADHGR 833
Qy 762 ----- 762
Db 834 GAIERNKGPEVASLPELRHSEDRFTPEPELOLRNLNENLGTNTVEELKDKLTISSSS 893
Qy 762 ----- 762
Db 894 DSIAMTPTIPSDKLAATEKTSGLSPRPNMSVHNHSLGTIVFQNNSSHLLQSGVPLESE 953
Qy 762 ----- 762
Db 954 EDNDSKILLEAPLMIQESSLRENVLSMESNRLEKKEERIRGPASLIDNALFKVNISVXT 1013
Qy 762 ----- 762
Db 1014 NRAPVNLTNKRTRVAIPULLIENSTSVWODIMLERNTEFEVTSLIHNTEMDRNTAL 1073
Qy 762 ----- 762
Db 1074 GLNHVSNKTTLSKNVEMAHQKREDPVPLRAENPDLSKIPFLPDMIKTHGKNSLSEQR 1133
Qy 762 ----- 762
Db 1134 PSEKQLTSLGSEKSVKQONELSEEKVVGDEFTKDELOEIPPNKSI:PFANLANYQEN 1193
Qy 762 ----- 762
Db 1194 DTYNOKKSPBEIERKEKLTQENVALPQAHMTIGTKNFKLNFLLSTKQVAGLEQPT 1253
Qy 762 ----- 762
Db 1254 PILQDTRSLNDSPHSEGIHMANFSKIREANLEGLQNTQWVERPSTTRMSSNASQHY 1313
Qy 762 ----- 762
Db 1314 ITQGRKSLKQPLSOGELIKFERKVIANDTSTQSKNMNYLAOGTTLQLEYNKEKRAIT 1373
Qy 762 ----- 762
Db 1374 QSPSLDSCSMNHVYTIQMNDSALPVAKESASPSVRHTDITLIPSOHNSHL:PASACNTFR 1433

Qy 762 ----- 762
Db 1434 ERTSGVQEGSHFLOEAKRNNLSLAFVTLGITEGOKFSSLSKSA:TNOPMKLENTVLLQ 1493
Qy 762 ----- 762
Db 1494 PGLSPDVKVELLSQVHVQDQSPFTKSNDSPGLDLMGKIFLOKTQGPVKMKTNSPG 1553
Qy 762 ----- 762
Db 1554 KVPFLKMAITESSEKIPSKLLGLVLAMDNHYDTQIPSEMKSO:KSQTNAFERKDTILPLG 1613
Qy 762 -----SONPVLRKHOREITRTTLOSQDEI 787
Db 1614 PCENNDSTAALNEGODKQORAMAKOGEPRCLQSQNPVPYKHNHOREITRTTLOPREDKF 1673
Qy 788 DYDDTISVEMKREDPDYDEDENSPNSFOKTRHYFLAVERLMDYGMSSPHVLANRA 847
Db 1674 EYDDTFSIEMKREDPDYGDYENOGLSFOKTRHYFLAVERLMDYGMSSPHILNRA 1733
Qy 848 QSGSVPOKKVYFOEFTDQSFOTPLRGELENEHGLIGPYTRA:VEENIMVTFERNOASRP 907
Db 1734 QSGDYQOKKVVFOEFTDQSFOTPLRGELENEHGLIGPYTRA:VEEDIMVTFERNOASRP 1793
Qy 908 YSFYSLSLSEEDQROGAEPKRNFKVNETKTYFMKVOHNA:APT KDEPDKAMAYFSDVD 967
Db 1794 XSFYSLSLSEDEDEQGAEPKRNFKVNETKTYFMKVOHNA:APT KDEPDKAMAYFSDVD 1853
Qy 968 LEKDVHSLIGPLVCHNTNTPAHAGROYVQOERLALFTTID:ERKSWFTENMRNCAP 1027
Db 1854 LEKDVHSLIGPLVCHNTNTPAHAGROYVQOERLALFTTID:ERKSWFTENMRNCAP 1913
Qy 1028 CNIOMEDPTEKENYFHAINGYIMDTPLGLYMAODQIR:YWL LMSGSNENHSHIFSGHV 1087
Db 1914 CNVOKEDPTLKENFRFHAINGYIMDTPLGLYMAODQIR:YWL LMSGSNENHSHIFSGHV 1973
Qy 1088 FTVRKKEEYKMAALNLYPGVEFTEVEMLP:SKAGIMRVECLIGEBHLHAGNSTLFLVYSNKCQ 1147
Db 1974 FTVRKKEEYKMAALNLYPGVEFTEVEMLP:SKAGIMRVECLIGEBHLHAGNSTLFLVYSNKCQ 2033
Qy 1148 TPLGASGHIDFQITAGQOGVAPKRLARLHYGSINAM:STKPEFSIKYDLAPMIIH 1207
Db 2034 TPLGASGHIDFQITAGQOGVAPKRLARLHYGSINAM:STKPEFSIKYDLAPMIIH 2093
Qy 1208 GIKTOGAROKFSSLYISOFIIMYSLDGKKQOTYRNS:GTILMVEFGNDSSGIKHINFP 1267
Db 2094 GIKTOGAROKFSSLYISOFIIMYSLDGKKQOTYRNS:GTILMVEFGNDSSGIKHINFP 2153
Qy 1268 PIIARIYIRLHPTHYSIRSTLRNEMGLDNLSCMP:LGMESKAISDAQITTASSYFTNMFAT 1327
Db 2154 PIIARIYIRLHPTHYSIRSTLRNEMGLDNLSCMP:LGMESKAISDAQITTASSYFTNMFAT 2213
Qy 1328 WSPSKARLHLOGRSNAMPPOVNPKEMLQYDFOKMKY:GTGVTQGVKSLTSMYKEFLI 1387
Db 2214 WSPSKARLHLOGRSNAMPPOVNPKEMLQYDFOKMKY:GTGVTQGVKSLTSMYKEFLI 2273
Qy 1388 SSSQDGHQWTLFFQNGKVVFOGNDSETPYVNS:LDPBLTRYLIRIHPOSWVHQTALRME 1447
Db 2274 SSSQDGHQWTLFFQNGKVVFOGNDSETPYVNS:LDPBLTRYLIRIHPOSWVHQTALRME 2333
Qy 1448 VLGCEAQ 1454
Db 2334 VLGCDTQ 2340

RESULT 2
ID 062730 PRELIMINARY; PRT: 2343 AA.
AC 062730;
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)

DE FACTOR VIII.
OS Canis familiaris (dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY, SPLEEN;
RA GORDY P.W., BOWEN R.A.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF049489; AAC05384.1;
DR PFAM: PF00394; Cu-oxidase; 3.
DR PFAM: PF00754; F5_P8_Type_C; 2.
DR PROSITE: PS01285; FA58C_1; 2.
DR PROSITE: PS01286; FA58C_2; 2.
SQ SEQUENCE 2343 AA; 265613 MM; D9C5A549 CRC32;

Query Match 80.9%; Score 6306; DB 6; Length 2343;
Best local Similarity 53.3%; Pred. No. 0;
Matches 125; Conservative 86; Mismatches 109; Indels 900; Gaps 4;

1 M Q I L S T C F L C L R F C S A T R R Y L G A V E L S M D Y M Q S D - L G E L P V D A R P P R P K S P F 59
1 M O V E L Y T C C F L C L P F S L S A T R K Y L G A V E L S M D Y M Q S D L S A L H A D T S F S R P G S L P L 60
1 N T S V Y V K K I L F V E T D L F N I A K R P P M G L G P T I Q A E V Y D V Y I L K N A S H P V S I H A 119
61 T T S Y T Y K R Y F E V E T D L F N I A K R P P M G L G P T I Q A E V Y D V Y I L K N A S H P V S I H A 120
120 V G S Y W K A S E A G E Y D Q T S O R E K E D K Y F P G S H T Y W Q V L K E N G P M A S D P L C I T Y S L S 179
121 V G S Y W K A S E A G E Y D Q T S O R E K E D D N Y I P G S H T Y W Q V L K E N G P M A S D P L C I T Y S L S 180
180 H V D L V K D L N S G L I G A L L Y C R G S L A K E R T Q T L A R F I L L F A V F D E G K S H S E T K N S L M D R 239
181 H V D L V K D L N S G L I G A L L Y C R G S L A K E R T Q T L O F E V L L F A F D E G K S H S T N S L I Q - - 239
240 D A A S A R A M P K M H Y N G Y V N S L P G L I G C H K S V Y W H V I G M T T E V H S I F L E G H T F L V R N 299
239 - - - - A E A Q H E L I N G Y V N S L P G L I G C H K S V Y W H V I G M T T E V H S I F L E G H T F L V R N 294
300 H R O A S L E I S P T F L T A O T L M D L G O F L L F C H I S S H O D G M A Y K V N S C P R E P O L R M K N N 359
295 H R O A S L E I S P T F L T A O T F L M D L G O F L F C H I P S H Q H D G M A Y K V N S C P R E P O L R M K N N 354
360 E E A D Y D D L D T S E M D Y V R F P D D N S P F I Q I R S Y A K K H P K T W H Y I A E E E D M Y A P L V L 419
355 E D - K D Y D D G L G S M D Y V S E D D S S P T O I R S Y A K K H P K T W H Y I A E E E D M Y A P S G P 413
420 A P D R S Y K S O Y L N N G P O R I G R K R Y K R E M A Y T D E F T K T R E A I O H E S G I L G P L Y G E V D T 479
414 T P N D R S H K N I L N N G P O R I G R K R Y K R E M A Y T D E F T K T R E A I O H E S G I L G P L Y G E V D T 473
480 L L I I F K N A S R P Y I Y P G I T D V R P L S R L P K G Y K H L K D P I L P G E I F K K T V Y E D G 539
474 L L I I F K N A S R P Y I Y P G I T V Y T P L H G R L P K G Y K H L K D P I L P G E I F K K T V Y E D G 533
540 P T K S D P R C I L T R Y S S F V M E R D L A S G L I G P L I C Y K E S V D O R G N O I M S D K R N V L I F S F D 599
534 P T K S D P R C I L T R Y S S F I L E N D L A S G L I G P L I C Y K E S V D O R G N O I M S D K R N V L I F S F D 593
600 E N R S W Y L T E N I O R F L P N P A G V O L D P E F O A S N I M S I N G Y F D S I O L S V C L H E V A Y W I L 659
594 E N R S W Y L T E D M Q R F L P N A D V O P H D P E F O L S N I M S I N G Y F D N L O L S V C L H E V A Y W I L 653
660 S I G A O T D E L S Y F F S G Y T F K K M Y E D T I L F P F S G E Y T F M S E N E N G M I L C H N S D F P N R 719
654 S V G A O T D E L S Y F F S G Y T F K K M Y E D T I L F P F S G E Y T F M S E N E N G M I L C H N S D F P N R 713
720 G M T A L L K Y S C D K N T G D Y E D S Y E D I S A Y L I S K N N A I E P R S F - - - - - 762
714 G M T A L L K Y S C N R N I D Y E D T Y E D I P P L L N E N N V I K P R F S Q N S R I P S T K E K O L A T T 773

QY 762 ----- 762
DB 774 T P E N D I E R I D L Q S G E R T Q L I K A O S V S S D L M L G O N P T P R G L F L S D L R E A T P R A D H S R 833
QY 762 ----- 762
DB 834 G A I E R N K G P P E V A S L R P E L H S E D R E T P P E L Q L R N E N L G T W T Y E L K K I D L K I S S S 893
QY 762 ----- 762
DB 894 D S L M T S P T I S D K L A A T E K T G S L G P P N M S V F N G H L G T I V F G N S S H L I O S G V P L E S E 953
QY 762 ----- 762
DB 954 E D N S K L L E A P L M N I O E S S L R E N V L S M E S N R L F K E R I R G P A S L I K D N A L F K N I S S V K T 1013
QY 762 ----- 762
DB 1014 N R A P V N L T N R K T V A I P T L I E N S T S W O D I M L E R N T E R K E V T S L I H N E T F M D R N T A L 1073
QY 762 ----- 762
DB 1074 G L N H V S K T I L S K N V E M A H O K E D P V P L A E N P D L S S K I P L P D W I K T H G K N S L S E O R 1133
QY 762 ----- 762
DB 1134 P S P K O L T S L G S E K S V K D O N F L S E K V V G E D E F T K T E L O E I F P P N K S I F A N L A N Y O E N 1193
QY 762 ----- 762
DB 1194 D T Y O E K S L E I E R K E K L Q E N V A L P O A H T M I G T K N F L K N L F L S T K O N V A G L E O P Y T 1233
QY 762 ----- 762
DB 1254 P I L D O T S L N D S P S E G I H M A N F S K I R E E A N L E G L N O T M O V E R P S T T R M S S A S O H V 1313
QY 762 ----- 762
DB 1314 I T O R G K R S L K O P R L S O E I K F E R K V I A N D T S O M S K N M N Y L A O G T L T O I E N E K E R A I T 1373
QY 762 ----- 762
DB 1374 Q S P L S D C S M R N H V T I O M N D S A L P V A K E S A S P S V R H T D L T I P S O H N S H L P A S A C N T F R 1433
QY 762 ----- 762
DB 1434 E R T S G V O E G S H F L O B A R N N I S L A F V L I G T E G G K S S L G K S A T N Q P M Y K K L E N T V L Q 1493
QY 762 ----- 762
DB 1494 P G L S E T S D K V E L L S Q V H V D O E D S E P T K T S N D S P G H L D M G K I F L O K T O P Y K M K N T S P G 1553
QY 762 ----- 762
DB 1554 K V P L K M A T E S S E K I P S K L L G V L A M D N H Y D T O I P S E E M S O K S Q T A F R K K D T I L P L G 1613
QY 762 ----- 762
DB 1614 P C E N N D S T A I N E G O D P Q R A M A K O G E P O R L C S Q N P P V S K H Q R E I T Y T T L P E E D K F 1673
QY 762 ----- 762
DB 1674 E Y D T F S I E M K R E D F I Y G D E D G L S F O K T R H Y F L A V E R L M D Y G M S R S P H I L N R A 1733
QY 848 O S G S V P O F K Y V P O E F T D S T O P L Y R G E L N E H G L L G P Y I R A V E N I N A W T F E N Q S R P 907
DB 1734 Q S G V O O F K A V F O E F T D S T O P L Y R G E L N E H G L L G P Y I R A V E N I N A W T F E N Q S R P 1793
QY 908 Y S F Y S S L I S E E D Q O G A E P R K N F V K P N E T K T Y F M K Y Q H H A P R K D E F D C A M A Y F S D V D 967
DB 1794 Y S F Y S S L I S I D E D Q G A E P R K R V N N E T K I Y T W K Y Q H H A P R K D E F D C A M A Y F S D V D 1853
QY 968 L E K D V H S G L I G P L L V C H T N T L N P A H G R O V T V Q E F A L F T I F D E T K S Y F T E N M E R N C R A P 1027

DB 1854 LEKVHSGILGILLICRSNTLPAHGRQVTVQEFALVFTIFDETSKWTETEMLEHNCRA 1913
QY 1028 CNIQMEDPFKFNRYFHAINGIMOTLPGLVNAOORIRWYLLSGSNENHSHIFPSHV 1087
DB 1914 CNOVEDPLTKENFRFHALNGYVKTLPGLVNAODOKAWYLLSMGSENHSHIFSGHV 1973
QY 1088 FVVRKKEEKALYNLYPGVFETVEMLPSKAGIMRVECELIQEHLAGMSTLEFLVSNKO 1147
DB 1974 FVVRKKEEKALYNLYPGVFETVEMLPSKAGIMRVECELIQEHLAGMSTLEFLVSNKO 2033
QY 1148 TPLGMAAGHIRDFQITASGOYQOMAPKLARLHSSGINAMSTKEPSPWIKVDLAPMII 1207
DB 2034 TPLGMAAGHIRDFQITASGOYQOMAPKLARLHSSGINAMSTKEPSPWIKVDLAPMII 2093
QY 1208 GIKTQAGARKFSSLYSOFIIMYSLDGKWKQYRGSTCTLWVFGNVDSGIKHNIENP 1267
DB 2094 GIKTQAGARKFSSLYSOFIIMYSLDGKWKQYRGSTCTLWVFGNVDSGIKHNIENP 2153
QY 1268 PIIRARYIRLPHYRSTIRLRLMELMCDLNSCMLPGMESKASDAQITASSYFTNMEAT 1327
DB 2154 PIIRARYIRLPHYRSTIRLRLMELMCDLNSCMLPGMESKASDAQITASSYFTNMEAT 2213
QY 1328 WSPSKARLHLQGRSNARPOVNNPKEMLOVDFOKTKVTVGTQGVKSLITSMYKEFLI 1387
DB 2214 WSPSKARLHLQGRSNARPOVNNPKEMLOVDFOKTKVTVGTQGVKSLITSMYKEFLI 2273
QY 1388 SSSQDGHOTLFFQNGKVVFOGNOSSFPVYNSLDPRLTLRLRHPSQWYHOTALRME 1447
DB 2274 SSSQDGHOTLFFQNGKVVFOGNOSSFPVYNSLDPRLTLRLRHPSQWYHOTALRME 2333
QY 1448 VLGCCEAO 1454
DB 2334 VLGCCEAO 2340

RESULT 3
043737 PRELIMINARY; PRT: 2224 AA.
ID 043737:
AC 043737:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, last annotation update)
DE FACTOR V.
OS Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
(1)
RP SEQUENCE FROM N.A.
RA BIRD C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: 299572; CAB16748.1; -.
DR PFAM: PF00394; Cu-oxidase; 3.
DR PFAM: PF00754; F5_F8_type_C; 2.
DR PROSITE: PS01285; FA58C_1; 2.
DR PROSITE: PS01286; FA58C_2; 2.
KW Blood coagulation.
SQ SEQUENCE 2224 AA; 251673 MW; 3BAAC1B7 CRC32;

Query Match 30.6%; Score 2387.5; DB 4; Length 2224;
Best Local Similarity 26.0%; Pred. No. 1.9e-172;
Matches 592; Conservative 278; Mismatches 483; Indels 923; Gaps 35;

QY 22 RRYVIGAVELSMYMODLDELVDARFPRVRPKSPFNFNSV-YKKTLEVEFTDHEFNI 80
DB 32 RQFYVAAGISWSRPE-----PTNSSLNLSTVSFKIVYRETEPT-FKK 75
QY 81 AKRRPMAAGLPTIOAEVDTVITLKNASHPVSLAVGVSYWKASEGAEYDDQTSOR 140
DB 76 EKQOSTISGLGLTVAEVDIIKVKHKNKADKPLSIHPIGIRSKLSSEGASVLDHTFPA 135
QY 141 EKEDDKVFPQGSHTTYWQVYLKENGPMASDPLCLTYLSLHVLDYKDLNSGLIGALLVCRE 200

DB 136 EKMDAAPGKREYIYEWMSIEDSGTPHDDPCLTHIYSHENLIEDNSGLIGLLCKK 195
QY 201 GSLAKETQ-TLHK-FILFAVPEDEKSMHSETPKNSIMQORDAASARAWKMTVNGVN 258
DB 196 GTLEGGTQKTFKQIYLLRLAVPEDESKMSOSS-----LMTVNGVN 239
QY 259 RSLDGLGCHKRSYVWVHVGNTTPEVHSIFLDSGHTFLVNRHROASLEISPTIFLAQTL 318
DB 240 GTMPDITVCAHDHISWHLGLGMSGPELFSIHFNQVLEONHHKVASATIVASSTANMT 299
QY 319 LMDGQPLFLPHISHOHODGEAYVYKDSCEPEPQLMKNNKEEDDDDLTSEMDVVR 378
DB 300 VGPEKWIISSLPLKHLQAGQAVIIDIKNCPKTRNLKKTRE----- 343
QY 379 FDDNNSPFIQIRSVAKKHPTVWYIAAEEDMDYAPLYLAPDRSKYQOLYNNNGPQRI 438
DB 343 -----QRRHAKRWEYFTAAEEVIMDYAPVIRPANNDRKYSQHLDNFSGOI 387
QY 439 GRKYKRVFMAVYDETEKTRAT--QHESGILGPLLYGEVDTLLIFKNQASRPYNIY 495
DB 388 GKHYKRVWYQYEDESE-TRKHTVNPNMKEDGILGPIIRAQVROTLKIVFKMASRPSYIY 446
QY 496 PHGIT-----DVRPLYSRLPKGVKHLKQEPILPGLFIFKRWVYVEDOPTKSDPCLT 549
DB 447 PHGVTFSEYEDENVSSFT---SGRNNTMIRAVQGEYTYKMWILPEDEPTENDAOCLT 502
QY 550 RYSSSFVMMERDLASGLGPLLICYKESVDORGNOIMSDKKNHVLFSYFENRSMYLTEN 609
DB 503 RPYSDVVDIMDIASGLIGLLICKSRSLDRQIORADIEQAVFVAFEDKMSMYLEDN 562
QY 610 IQRLPNPAGVQLEDPEFOASNIHSHINGVYFDSL-QLSVCLHEVANYIILSTAQIDFL 668
DB 563 INKFCENDEVKRDPDKPYEESINISTINGVYESITTLGFCFDPTVOWHFCSVGTONEIL 622
QY 669 SYFESGTYFKHKMYEDTLTLPFSGETVEMSMENPGMILIGCHNS----- 715
DB 623 TIFHTGSHFIYGRHEDTLTLPKMGESVYVMDNVGTMLTSMNSPSRKKLAKPRDV 682
QY 715 -----DFRNR----- 720
DB 683 KCIPIDDDSYEIEFPPESTVMAATPKMMDRLPEDESEADYDQNRLLAALGIRSFNS 742
QY 720 -----GMTAL----- 725
DB 743 SLNDEEENLALALENGTEFVSSNTDIIYGSNYSPPSNISKEFTVNNLAEPOKAPSHQ 802
QY 725 -----LAVSODKMTGDYEDSYED----- 745
DB 803 ATTAGSPRLHLIGKNSVLSNSTAEHSSPYSEDPIDPLQPDVYTGIRLSLAGGEFSQEH 862
QY 745 ----- 745
DB 863 AKHKGPYERDQAAKHRSWMLLAHKYGRHLSDQTGSPSGMRPWEDELPSQDGTSPSMR 922
QY 745 -----ISAYLSKNN----- 756
DB 923 PKMDPPSLLILKQSNSSKILVGRWHLASEKSYEIIQDTDEDTAIVNMWLSIPQNASRAW 982
QY 756 ----- 756
DB 983 GSTPLANKPGQSGHPKPPVRRHKSLOYRQDGGSKRLKKSOFILTKKKKKKTHHAP 1042
QY 756 IPRSP-----SQ 763
DB 1043 LSPRTFHLPLKSAVNTFESRRLKSLVYLKHSNETSLPTDLNQTLPMSDFGWIASLPDHQ 1102
QY 764 N----- 765
DB 1103 NSNDTQASCEPGLYQVTPREHYQTEPIODPDQMHSTSDPSHRSSSPELSEMLEYDRS 1162
QY 765 -----PVLKRR----- 772

Dh	1163	HKSPETDISQSPSESEHEWQVYI	ISPDLSQVTLSPDLISQTNLSPDLSHHTLSPDLIQNL	1222
QY	772	-----	-----	772
Dh	1223	SPALGOMPISPDLSHTLTLSPDL	SHHTLTLSPDLISQTNLSPDLISQTNLSPALGOMPLSPDL	1282
QY	772	-----	-----	772
Dh	1283	TTLSLDFSQTNLSPDLSEHMTLS	PELSQTNLSPALGOMPISPDLSHTLTLSPDLISQTNLSPE	1342
QY	772	-----	-----	788
Dh	1343	LSQTNLSPALGOMPLSPDP	SHHTLTLSPDLISQTNLSPDLISQTNLSPDLSEKPLADLSQPL	1402
QY	788	-----	-----	788
Dh	1403	TPDLIDQMTLSPDIGETDLS	SPNFGOMLSPDLSQVTLSPDISDPTLLPDLISQISPEPDDQ	1462
QY	788	-----	-----	792
Dh	1463	IEYPSSESSQSLLLQEFNESE	FPYDLOGMPSPTLNDTFLSKENPLIVIGLSKDGNDY	1522
QY	793	ISYEMRKE-----	-DFDIYDE-----DENOSP-----RSFOKTRHY	823
Dh	1523	IEIFPKREVOSSEDDVAIED	YVYDDPYKTDAVTNINSRDPDNTAAMTILSNKNNKRY	1582
QY	824	FIAAVERLMDYGNSSSPHYLRN	R--AOSGSVPD--FKRYVFOEFTDGSFTOPLRYGELN	878
Dh	1583	YIAAEEISMDY-----	SEFYQRETDIEDSDIPEDTYYKRYVFRKYLDSTFKRDPREYE	1638
QY	879	EHGGLLPYIRAEVDENIM	TFPNOASRPYSFSSLSISE-----EDQOGAEPPKN	930
Dh	1639	EHGGLLPYIRAEVDVIOY	FRFNKLSRPSLNAHGLSYEKSSEGTYESDDEPEWKEBDN	1698
QY	931	FVKFNFEKTYFMKQOHMA	PTKDECFCKAMAYSDVDLEKDYHSGLIGPLVCHINTPLN	990
Dh	1699	AVQGNSSYTYVMHATERS	GEPSGSACRAWAYISAVNPEKDHSGLIGPLVLCQGLIHK	1758
QY	991	AHGHOVTVOEFALFTT	IFDETKSMYFTENNERNCRAPCNIOMEDPFFKXNYFFHAINCYI	1050
Dh	1759	DSNNPMDMREVFLL	FMFDEKSKMYTEKKSRSRSMR-----LTSSEMKSHFHA	1812
QY	1051	MDTIPGLVMAQDORIK	MYTLISMGSENENIHSIHPSGHVFTVRKKEEYKMAIYLVPGVEFT	1110
Dh	1813	Y-SLPGIKMYEOWEVR	LHLNLINIGSODIHVHFHGOPLLENGKOHQOLGWPMLDGPSEFT	1871
QY	1111	VEMLPISAGIWRVBC	LGEHLGELHAGMSTLFLYENKQOTPLGMASSGIRPOTIATAGQGYQ	1170
Dh	1872	LEMKASPGWMLL	TEGEMORAGMOPFLINDRCDRMFGSLGTGISOSQIKAEFPGY	1931
QY	1171	WAPKLARLHSGSIN	AWSTRE---PES---WIKVDCLAMITHIGITQCAROKFSSLYIS	1224
Dh	1932	MEPLRLALNNGGS	VNANMSYERKLAIEFASKRWIOYDMQKEVITITIGIOTQAKHYLASCTTY	1991
QY	1225	QFTIMSLDGGKWO	TYRGNSTGLTAMVFEGVNDSSGIRKHNIFNPPIIARYIRLHPHYISR	1284
Dh	1992	EFYVAYSNNOI	NMQIFKGNSTRMVFNGNSDASTIKENQDPPIYARVIRISPPRAYNR	2051
QY	1285	STLMEMLMGD	LNSCMPRLMESKASISDAOITASSFTTNMFRAT--WPSARALHLOGRNSA	1343
Dh	2052	PTLRLLELLOGE	EVNGCSPTPLDMENGKIEKNKOTIASSFRKSMWGGYVWPFARLNAAGRYNA	2111
QY	1344	WRPOVNNPKEM	LQVDFQCTKRYGVYGTTOGKSLTJEMYKREFLISSQODGHOMTLFEPQNG	1403
Dh	2112	WQAAANNKQW	LEIDLKIKITAITOIGCKSLISEMYKASTYIIHISEGVGVMAKRYRLKS	2171
QY	1404	KV--KVFQGNDS	FTPYVNSLDPPLRLYRLIHPQSWVHQIALRMEVLGCEAODLY	1457
Dh	2172	SMVDKIFEGNT	NTKGHYKFNFPPIISRFRIVRIPKTMWNSIALRLELPGFC---DIY	2224
RESULT	4			
	0887783			

```

ID      088783          PRELIMINARY:          PRI: 2183 AA.
AC      088783:
DT      01-NOV-1998 (TREMBLrel. 08, Created)
DT      01-NOV-1998 (TREMBLrel. 08, last sequence update)
DT      01-NOV-1999 (TREMBLrel. 10, last annotation update)
DE      MURINE COAGULATION FACTOR V.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC      Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN      [1]
RP      SEQUENCE FROM N.A.
RA      YANG T.L., CUI J., REHOMTULLA A., MOUSSALLI M., KAUFMAN R.J.,
RA      GINSBURG D.;
RT      "The structure and function of murine factor V and its inactivation
RL      by protein C.;"
RL      Blood 91:0-0(0012).
[2]
RP      SEQUENCE FROM N.A.
RA      GINSBURG D., YANG T.L., CUI J., YANG A.;
RA      Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.
DR      EMBL; U52925; AAC9553.1; -.
DR      PFM; PF00394; Cu-oxidase; 3.
DR      PFM; PF00754; P5_F8_type.C; 2.
DR      PROSITE; PS01285; FA58C_1; 2.
DR      PROSITE; PS01286; FA58C_2; 2.
SQ      SEQUENCE 2183 AA; 247228 MW; 5F22D6A CRC32;

Query Match          30.5%; Score 2374.5; DB 11; Length 2183;
Best Local Similarity 25.7%; Pred. No. 1.8e-171;
Matches 582; Conservative 291; Mismatches 491; Indels 899; Gaps 33.

QY      8 CFFCLCL--REC-----FSATRRYLGAVELSMWYMOSELGELPVDARFPRVPERK 56
      |||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      7 CFFLLVLTGTRMAGSGHQAEALQRLQRYVAAGILMYNHP-----PRD 51

QY      57 EPENTSVYKKTLEVEFDHLEFNIAKPRPPWMLGPTIQAQEVYDTVYITLKNMASHVPS 116
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      52 PSLNISPFKKIVREY -EQYFKKKPRSSNGSLGLPTLAEVGVDIRKHFNRKADKPLS 110

QY      117 LHAQVSVYWKASSEGAEYDDQTSQRKEDDKYPPGSGSHYVWQYLKENGPMASDPLCLTYS 176
      : : : : : |||| : : : : : : : : : : : : : : : : : : : : : : :
DB      111 IHPQIKYKSESEGSYADHTFPARKKDDAVAPEGELYEWIVSEDSQTPDPDLCTHI 170

QY      177 YLSHVDLVKNDLSIGLALVLCREGSLAKKERTQT -HKFILLFAVPEGKSMHSETKNS 234
      : : : : : |||| : : : : : : : : : : : : : : : : : : : : : : :
DB      171 YSYENLVODPFSGLIGPLILCKKSTLVEDGTQKFKDKOHVLEAFVDESK----- 223

QY      235 LMQDDASASARAMPKMTVNGVYNRSLPGLIGCHRRKSYVMHYIGKTPPEVASIFLEGHT 294
      : : : : : |||| : : : : : : : : : : : : : : : : : : : : : : :
DB      223 -----RSQSPSLMTYINGFVNKTPMDITVCADHYSWHLIGNSSGELFSLHFNQV 274

QY      295 FLVRRHQASLEISITTLTFLAOTLMDLGGFLFGLFHISHQHDGEAYKYKDSCEEPOL 354
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      275 LEONQHKYSTVLVSATSTTAMTMSPEGRWVSSLIRKHYAQAGQAVYIDINKCKKTR- 334

QY      355 RMKNNHEAEADYDDLLDSEMDVYRRDDONSPEFIQIRSYAKKHKPTWVHYIAAEDMDY 414
      : : : : : |||| : : : : : : : : : : : : : : : : : : : : : : :
DB      334 -----SPK--TLTRDQRKYMKEVFLIAAEVIMV 362

QY      415 APLVLPADDRSKSYOLNNGPQIRGKKKYAFMAVYDET -KTRREALQHESSGLGPL 472
      || : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      363 APLVIRAMNDKIRSHOHDFNSMGIKHKKVYRQYEETFLKRDNSIKSGSLGPLVI 422

QY      473 YGEVDDTLITPKNOASRPYNIYPHGITDVRPLYSRRLPKGV-----KILKDPILPG 525
      : : : : : |||| : : : : : : : : : : : : : : : : : : : : : : :
DB      423 RAQVQDTLKIYKKNASRPYSIYPHGVT-----PSPYEDGINSSTSGSHITIRPVQG 476

QY      526 ELFKKMYTVYEDGTSKSPCLTXYYSFVMMEDLASGLGPLILCYKESVDQRGNDI 585
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      477 EFTFKWNLLEFDETEQDLQCLTPPYSDVVTDLASGLGLLILKSRSLDRGVQR 536

586 MSDKRNVLLEVFEDNRSVLYTENQRELPJNPAGVQLDEPPEQASINMHSINGVYEDSLQ 645

```

```

Db 537 VADIDQAVFANFENDKSNYIEDNINKFCENDEYKRDPRKYEENKSTINGVPEST 596
Qy 646 -LSVCLHEAVYVYILSIGAQTDFLSVFESGYTFKKHMYEDTLPPSGETVEMSMENP 704
Db 597 TLGCFDDTVQMHFCSVGHHDILTFHGFHSIFYGRHEDTLTFPMGESVTVTMNV 656
Qy 705 GIMTICHSDFRNGMTALLKVSQCDKNTGY-YEDSYE----- 744
Db 657 GTWMLTTMNSNKRRLRLRFEDVCKNR--DYDNEDSTEIYEPAPYSMTTRRHDSLE 713
Qy 744 -----DISAYLLS-----KNAIEPR----- 760
Db 714 NEFGIDNEDDYOYLASSLGIRFSNSSLNPEENFNLTALALENSSEFISPTDRYVD 773
Qy 760 ----- 760
Db 774 SNSSRLSKINNKLDFORTLPFGGATVAGTLRLNLIGLDENFVLSSTEHRSSSYHEN 833
Qy 760 ----- 760
Db 834 DMENPOSNTMYLLPLGPKSGNREODXPKTIKTRPHMMHRSKMAKPAKIGRHSN 893
Qy 760 ----- 760
Db 894 PKNSYSGMSEEDIPSELLPLKQKITSKFLNRMRVASEKSEYELIANGEDTDVXLIN 953
Qy 760 ----- 760
Db 954 SPONQITVPRGESHNTTRKPSDLPTFSGVGHKSPHYROEENSGFQROLFTRTK 1013
Qy 760 ----- 760
Db 1014 KKKNKLLALHSPSPGFPBLRGHNHSPPRDLRLHSLHKSNETALSPDLNQTSPEN 1073
Qy 760 -----SFQONPP 766
Db 1074 STDRSLPDYNOYKNDTEQWSSSLDYQSVPAEHSPTPAODPDQTHSTTDPSTYRSP 1133
Qy 767 VLKR----- 771
Db 1134 ELSQGLDYLSHFPDDIGLTFPPDQOKSSFSSDDQAISSDLSLFTISPELDQTI 1193
Qy 771 ----- 771
Db 1194 IYPDQLLSPDNQKTSPPDLGOVPLSPDNQKTSPPDLGOVPLSPDNQKTSPPDLG 1253
Qy 771 -----HQR----- 774
Db 1254 QVPLSLDDNQKTSPPDLGOVPLSPDNQKTSPPDLGOVPLSPDLGOVPLFPE 1313
Qy 774 -----EITRTLQSPQOE----- 787
Db 1314 DNQNYFLDLSQVPLSQNQBETSTDLTLSPDEGQVPLSPDLQPLPSPDNQVTVSPD 1373
Qy 787 -----IDVDDN----- 793
Db 1434 LPELNLRTLPHPLDLTHPPSPSPPTLNTSLSRKFNPLVVVGLSKVDDDDVEIYSEPER 1493
Qy 793 ISVEMKKEDFDIYDEDESNOSRPFQKTRH-----YFIAVERLMDY 834
Db 1494 IDEYIADDDVTVNDPYRTJTRDVNSSRNPDITAAWYLGHHGKKFYIAAEIWNV 1553
Qy 835 GMSSPHYLNRRAQSGVPO---FKYVFOETDGSFTQPLRGELNEHGLGAPYIRAE 891
Db 1554 AEFQASEM--DHEDTGHTPKTYTKKVVFRKYLDSTFTSRDPAEYEHGILGIPVIRAE 1611
Qy 892 VEDNIWTFPNOASRPYSFSSLSIYE-----EOROGAEPRKKFVKNKTYEWM 943

```

```

Db 1612 VDDYIOVFKMLASRPYSLAHGLSYEKSEKTYEDESEPEWEOEDDAVOVONSSYTVMH 1671
Qy 944 VOHMAPTKDEDFCKAMAYSDVDLEKDVHSGLIPLYOCHTNTLNPDAHGOVYQEFAL 1003
Db 1672 ATKRSQPNPGSACRAMAYSAVAVENDIHSGLIGPLLCRKGLIHERNLPHDMREFVL 1731
Qy 1004 FFTIETDKSWYFTEENBERNCRAPCNIOMEDPPEKERYRPHALNGYIMDTLPGLVMAOQ 1063
Db 1732 LFMVFEKKSMTYKES--KGSR-----RISPEEKNHAKFYAINGMITYN-LPGIRMYEOE 1783
Qy 1064 RIRMTLSMGSNENHSHSGHYFYRKKREYTMALXNLYPGVFEEVEMLPKSAAGIMRV 1123
Db 1784 WVRHLMLMGSSRDIIHVHFQOTLNDNRKQHOLGWAPLLPGSEFKLEMKASPGWMLL 1843
Qy 1124 ECLIGEHLHAGMSTLFLYVSKCOTPLGMASGHTRDQITASGQYGOVMAPEKLARLHSGS 1183
Db 1844 DTEVENOVAGMOTPELILDECKPMGLSTGYISDQIKASELYLTWEPRLARLNNAGS 1903
Qy 1184 INAMSTKE----PFS-WIKYDLAPMIHIGIKTOGAROKFSSLYISOFTIMYSLDGKKW 1237
Db 1904 YNAMSIEKTALDPFIKPIWIOVDMQKEVVYTGIOGAKHYLKSCFTEFQVAYSDDQTNW 1963
Qy 1238 OTYRANSIGTLMVFFGNDSSGICKHNIFNPPIIARYIRLHPTHTSISLMEIAGDOLN 1297
Db 1964 QIFRGSKGSVMYFTGNDSGTIKENRDLDPPIVARYIRIHPTKSYNRPTLLELOGCEN 2023
Qy 1298 SC5MPLMESKASIDQATASSYFTNMFAT--WSPKARLHLGSRSNMARPQVNPKEWLO 1356
Db 2024 GCSITGLIEDGRIDQKQITASSFKKSWMGDYWPESLALNAGVNMAMQAAANNKQLO 2083
Qy 1357 VDFQTKMVTGVTQGVKSLTSMYKEFLISSODGHQWTLFQNGKV--KYFQGNDS 1414
Db 2084 VDLKIKRYTALVTOGCKSLSEMYKSYSTIO5POGVAKMPYRQKSSVMYKIFEGNSNT 2143
Qy 1415 FPPVYNSIDPPLFTRYLHPQSWVHOIALMEVYLGEADOLY 1457
Db 2144 KGHMKNFNPPLTSKFRILIRKTMOSTALRLLEGC---DIY 2183

```

RESULT 5

Q92024 PRELIMINARY: PRT: 1157 AA.

AC Q92024.

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DE 01-MAY-1999 (TREMBLrel. 10, Last annotation update)

HEPHESTIN.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-C57BL/6J;

RA VULPE C.D., KTO Y.M., MURPHY T.L., COWLEY L., ASKWITH C., LIBINA N.,

RA GITSCHER J., ANDERSON G.J.;

RT "Hephaestin, a ceruloplasmin homologue implicated in intestinal iron

transport, is defective in the sla mouse.";

RL Nat. Genet. 21:195-199(1999).

DR EMBL: AF082567; AAD16035.1;

DR PROSITE: PS00079; MULTICOOPER.OXIDASE1; 3.

DR PROSITE: PS00080; MULTICOOPER.OXIDASE2; 1.

SEQUENCE 1157 AA: 129682 MW: 10083457 CMC32;

Query Match 23.2%; Score 1808; DB 11; Length 1157;

Best local Similarity 34.2%; Pred. No. 8.4e-129;

Matches 403; Conservative 171; Mismatches 417; Indels 186; Gaps 20;

```

Qy 20 AFRRYLGAVELSWDIQMSDLBELPYDAFPPRPVPSKSF-----PENTSVYKKTLYVEFTD 75
Db 24 AIRNYLIQIDQMNYAPAGRNVITNQTLLNDTVASSFLSKGNRIRIGSSYKKTIVYKEYSD 83

```

QY 76 HLFNIAKPRPMGLGFTIOAEYVDVTVITLKINASHPVLSHAAGVSYMKASGAEYD 135
 DB 84 GTYIEELAKPAMLGFTIOAEYVDVTVITLKINASHPVLSHAAGVSYMKASGAEYD 143
 QY 136 QTSQREKEDKVPFGSGHTYVWVLKENGPMASDPLCTYSYSLVHDVLYKDLNSGLIGAL 195
 DB 144 GSSGYLAKADSDVPFGSGHTYVWVLKENGPMASDPLCTYSYSLVHDVLYKDLNSGLIGAL 203
 QY 196 LVCEGSL-----AKEKOTLHKFLFLFAVEDEGSMSE-----TKSLMODDASAR 245
 DB 204 ITCKRGTLGDSNPPORRVDVDFNLFPSVIDENLSWHLDDIATYCSPPASVDEDEGDFQ 263
 QY 246 AMPKHTVNGVYNSRLPGLICHRKSVYWHYIGKTPTEVHSIFLEGTFFVNRHROASL 305
 DB 264 DSNMHAINGVFNGLPGLICHRKSVYWHYIGKTPTEVHSIFLEGTFFVNRHROASL 323
 QY 306 EISPTFLTAOTLMDLGOFLFLFCHISSHQHDEAVYKVDSCPEEPLRMKNNEADY 365
 DB 324 NIPFATVTAAMVPQKSGTWLISCEVNSHLSRGMQAFYKVDSCMDPPV----- 373
 QY 366 DDGLDSEMDVYRRDDNSPSFIQIRSAKHKPTWYHIAAEEEDMDYALVLPDRS 425
 DB 373 -DOLTG-----KVROFTEAHETQWYGP--IGYDGT 402
 QY 426 YKS-----QYLNNGPORIGRKYKVRPMAYTDEFKTRREALIOMES--GILGPLYG 474
 DB 403 GKSLREPGSGDKFYOKSSRIIGTYKVRKYEAFODEFQRYHDEETHGILGPLYRA 462
 QY 475 EVGTLTIFERNQASRPYNYPHGITDVRPLYSRLPRGVKH---LKDPIPLPEIFK 531
 DB 463 EVGTLTIOVFNRAISOPEFSIQPHGVFEYKNEGTVYNGTSHPRKAKSF-----EKVY 517
 QY 532 WTVYVEDEGPTKSDPRCLTRYSSFYNNMRDLASGLIGLFLCYKESVDORGNOIMSKRN 591
 DB 518 WTVYVPHAGPTAADPACLTWMTFSADPRDNTSGLYGVLCKAGALADKOKGVKEF 577
 QY 592 VILSVFEDENSWYLTENIQRFLEPNAG-----VLEDEPE--FOASNIMHSINGVYFDSL- 645
 DB 578 FLTLTVDENSKWYNNAN-----QAAGMLDSRLSEVEGFODSNRHHAINGFLFSMLP 631
 QY 645 QLSVCLHEVAWYVILSIGAOTDPLSVFESGYTFKHKMYEDTLTLFPESGTEVMSMNP 704
 DB 632 RLDMCKGDTYAMHLLGTEITDVHGVMEGNTYVLOGKRGKGMVLPFTYTAIMQDPN 691
 QY 705 GLMILGCHNSDFRNRGMTALTKVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRPSQN 764
 DB 692 GIFEIYCOAGSHREGMALYNSQC-----SSHOD----- 723
 QY 765 PVLKRRHQREITRTTLGSDQDEIDYDITISVEMKKEDFDIODEBENOSPRSFQKTRHYF 824
 DB 723 -----SPRHYOASRY 735
 QY 825 IAAVERLMDY-----GMSPPVLRNRASGVPOFKKVVFOEFTDGSFTQ 870
 DB 736 IMABEIMDYCPDRSWELEHMTSEKSYGHVFLSNKGLGSKKAKVFEYIDGTFR 795
 QY 871 PLYGELEHNLGLGPTIRAEVDINMTFRNQASRPYFYS--SLISYEDORQGAERK 929
 DB 796 PRPSGGEENHIGLIPRLIRGEVGDILTVFKNKASRPYSIAHAGLESNTGGPQAAE--- 853
 QY 930 NFVAPNKKYFWYQNHMAPTKDEFOCKAAYPSDVULEKDVHSGLLGPLLVCHTNTLN 989
 DB 853 ---PGEVLYQWNPERSGPPSDSACVSWIYSAVPIKDMTSGLVPLVICNGILE 908
 QY 990 PAHGRVYVQEFALFTTIFDETKSWYFENNER--NCRAPCNIOMDPFKENYREHAIN 1048
 DB 909 PNGRNDMDREFALLFLTFEDENOSWYTKENATATYQPOSSHVNLKDAFLFSLSKNHAING 968
 QY 1049 YIMDTLPGLVMAODORIMWYLLSGNSNHSIHSGHVFYVRKKEEKMALNYLPGVF 1108
 DB 969 KLYANLRLITVYQGERVAMYLAMQODDITVHHAESFYUNQOSYRADVVDLPGTF 1028
 QY 1109 ETVENLPSKAGIMWRECLIGELHLAGMSTLFLVYSNK 1145

DB 1029 EYEMVANSNGETWMLMCHYTDVHAGMETIFTVLSHE 1065
 RESULT 6
 ID 075180 PRELIMINARY; PRT; 891 AA.
 AC 075180
 DT 01-NOV-1998 (TEMBREL. 08, Created)
 DT 01-NOV-1998 (TEMBREL. 08, Last sequence update)
 DT 01-MAY-1999 (TEMBREL. 10, Last annotation update)
 DE KIA00698 PROTEIN.
 GN KIA00698.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN.
 RX MEDLINE; 98403880.
 RA ISHIKAWA K., NAGASE T., SUYAMA M., MIYAJIMA N., TANAKA A., KOTANI H.,
 RA NOMURA N., OHARA O.;
 RT "Prediction of the coding sequences of unidentified human genes. X.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro."
 RL DNA Res. 5:169-176(1998).
 DR EMBL; AB014598; BAA31673.1; -.
 DR FPM; AF00394; Cu-oxidase; 3.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
 DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
 SQ SEQUENCE 891 AA; 100624 MW; 4EA30F16 CRC32;
 Query Match 17.2%; Score 1341; DB 4; Length 891;
 Best Local Similarity 31.1%; Pred. No. 1.9e-93;
 Matches 321; Conservative 148; Mismatches 358; Indels 206; Gaps 23;
 QY 250 MHTVNGVYNSRLPGLICHRKSVYWHYIGKTPTEVHSIFLEGTFFVNRHROASL 309
 DB 1 MHAINGVFNGLPGLICHRKSVYWHYIGKTPTEVHSIFLEGTFFVNRHROASL 60
 QY 310 ITFLTAOTLMDLGOFLFLFCHISSHQHDEAVYKVDSCPEEPLRMKNNEADYD 369
 DB 61 ATFTAEVWPEPCTWILSCOVNSHFRDGMALTKVNSCANPV-----DL 108
 QY 370 TDEMDVYRRDDNSPSFIQIRSAKHKPTWYHIAAEEEDMDYALVLPDRS 420
 DB 109 TG-----KVROFTEAHETQWYGP--IGYDGT 402
 QY 421 PDDRSYISQYLNNGPORIGRKYKVRPMAYTDEFKTRREALIOMES--GILGPLYG 478
 DB 142 PG--SISDKFEOKSSRIGTYKVRKYEAFODEFQRYHDEETHGILGPLYRA 462
 QY 479 TLLIIFKNQASRPYNYPHGITDVRPLYSRLPRGVKHLPKLPGLPEIFK-----YK 532
 DB 200 TIQVFTNRASOPSMOHV-----FTEKDYESTV--YNDGSSYPLGLVAKPFEKVTY 252
 QY 533 TVTVEDEGPTKSDPRCLTRYSSFYNNMRDLASGLIGLFLCYKESVDORGNOIMSKRN 592
 DB 253 TVPPHAGPTAADPACLTWMTFSADPRDNTSGLYGVLCKAGALADKOKGVKEF 577
 QY 593 ILFVFDENSWYLTENIQRFLEPNAGVQ-----EDPE--FOASNIMHSINGVYFDSL- 645
 DB 313 LLFTVLDEKNSWYNNAN-----QAAGMLDSRLSEVEGFODSNRHHAINGFLFSMLP 631
 QY 645 QLSVCLHEVAWYVILSIGAOTDPLSVFESGYTFKHKMYEDTLTLFPESGTEVMSMNP 704
 DB 366 RLDMCKGDTYAMHLLGTEITDVHGVMEGNTYVLOGKRGKGMVLPFTYTAIMQDPN 691
 QY 705 GLMILGCHNSDFRNRGMTALTKVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRPSQN 764
 DB 426 GTFEYCOAGSHREGMALYNSQC-----SSHOD----- 723

```

QY 765 PVLKRNHOREITRTLLQSDQEEIDYDDITISVEMKKEDFDIYDEDENSPFSOKTTRHYE 824
Db 452 ----PGHQ-----ADPROXYQARITYY 469
QY 825 IAAVERLMDYGMSSS-PHYLRNRAOSGSV-----POFKVVEQFETDGSFTQ 870
Db 470 IAAEEVMDYCPDRSWEHMHNOSEKDSYGIIFLSNKDGLGSRKKAKAFREYTDGTERI 529
QY 871 PLXRGELNHLILGPYIAEVEDNINMVFERNQASRPYSFS-SLISYEDDROGAEPKX 929
Db 530 PPRRGPEEHLILGPYIAEVEDNINMVFERNQASRPYSFS-SLISYEDDROGAEPKX 929
QY 930 NEVKNETRYFWKYOHNAFPTKDEFCKAMAYFSDVDLEKDVHSGILGPLYCHTNTLN 989
Db 587 ----PGEVVTQWNIPIERSGPGNSDCAVSWIYSAVDIKWYSGLVOPLAICQIGILE 642
QY 990 PAHGRQVYQEAFLFTTIDETKSWFTENM-ERNCRACNIONMEPTTKENTRFAIAG 1048
Db 643 PHGRSDMDREFALFLIFDENKSWYLEENVAITHGSQDGSINLODETELESNKMAIAG 702
QY 1049 YIMDTLPGIYMAQDORIRMYLILSMGSNENIHISFGHYFTVAKKDEYKMAIYNLPYGF 1108
Db 703 KLYANIRGLMTQGEVAMITLMAQDVDTHTFHASEFLRNGENYRADVDLFPCTF 762
QY 1109 ETVEMLPSKAGIWRVBCILIGEHLHAQMSLEFLVYSNKCOTPLIGMASGHIIDFOITASGOY 1168
Db 763 EYEVEMASNPCTWLMHCHTDHVAHAGMETLFTVFSR-----TEHLSPLTVIKETE 813
QY 1169 GOMAPLARLHSGSINANSTKEPFSIKYVDLAPMIHIGITOGAROKFSSIIYIOFII 1228
Db 814 KAVPPDIE--EGNVKMLMQIPK--NVEMLASVLY-----ALSTVILLV 855
QY 1229 MYSLDGKKMOTYR 1241
Db 856 VYALGGVWYQHR 868

RESULT 7
075659 PRELIMINARY; PRT; 782 AA.
AC 075659;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
DE 01-MAY-1999 (TREMBLrel. 10, last annotation update)
DJ 64618.1 (COAGULATION FACTOR V
(ACTIVATED PROTEIN C COFACTOR), COAGULATION FACTOR VIII
(ER 1.16.3.1, FERROXIDASE) LIKE) (FRAGMENT).
GN DJ46618.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA WILKINSON J.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL030998; CA19742.1;
DR PFAM; PF00394; Cu-oxidase; 2.
FT NON_TER 1
FT NON_TER 782
SQ SEQUENCE 782 AA; 87939 MW; 7DD1IE52 CRC32;

Query Match 15.8%; Score 1230; DB 4; Length 782;
Best Local Similarity 32.8%; Pred. No. 4,3e-85;
Matches 293; Conservative 121; Mismatches 295; Indels 184; Gaps 18;
QY 65 YKKTLEVEFTDHLFNIAKPRPMGGLGPTIQAEDVTVTILKNMASHVSLHAGVSY 124
Db 13 YKTIYKEVYKDDSYTDEVAQPAWLGFLGPAVEGVIIILHKNFATRPYTIHPHGVFY 72
QY 125 WKASEGAEYDQTSQREKEDKVPFGSGHRYVWQVLKENGMSADPLCLTYSYSHVDLV 184

```

```

Db 73 EKDESGSLYPDSSGPKADDSVPPGSGSHIYNMTIPEGHAPTDADPACLTWIIHSHVAP 132
QY 185 KDLSGLIGALLVCREGSLAKETQTLHKFFILFAVDEGKSMH-----SETKNSIMOD 238
Db 133 KDITGLIGPITCKRG-----TDHDFLLFSVDENSHMLNINMNTFYCDPASVD 184
QY 239 RDAASARAMPKMTVNGVYNSRLPGLIGCHKSVYMWVHVGITTEPEVHSIFLGHTLYR 298
Db 185 KEDETFOESNMHAIINGVEFNGLPMLNCAQKRYAMHIFGNGEIDVHTAFHGMULTR 244
QY 299 NHRQASLEISPTLLTQTLMDLGOFLFCHISHOHDGEAAVYKVDSCPEEQLRMKN 358
Db 245 GHHTDVANIFPATTVYLAEMVWEPGTWILSCQVNSHRDQKQALIKYKSCMAPPV----- 301
QY 359 NEAEADYDDDLTSEMENVYRDDNDSPSFIQIRSAVKHNPRTWNYIAAEEDMDYAPL- 418
Db 301 ----DLITG-----KVROYFLFAHRIQWDYGPBG 325
QY 418 ----VLAPDRSYKSQYLLNNGPQIRIGKRYKKVPMATYDTEFKTREALIOHES--GI 467
Db 326 HDGSTGKNIRBPG--SISDKFFOKSSRIIGTYKVRLEAFODETFQKMLEDRHLGI 383
QY 468 IGPILYGEVDTLLIFKNOASRPYNTIYPHGITDVRPLYSRLPKGVYHLKDPILPGEI 527
Db 364 IGPVIRAEVGDITQVYFYNRASQSPFSMOPHGV-----FYEDYEGTV--YNDGSSYPOLY 436
QY 528 FK-----YKWTVEVEDPRTKSDPCLTRYYSFVNMERDLASGLIGPLICYESVDOR 581
Db 437 AKPEKATYRMTVPRHAPTAQODPACLTWMTYFSADPFRDINSGLVGLVLCRAGALCAD 496
QY 582 GNOISDKRNYLSEVEDENSWITENIQFLPRPAGVQ-----BDPE-FOASNLH 634
Db 497 GKQKGVKDEFFLLFTVLDENKSWYSN-----AQQAAMDFLLSDDIGFQDSNNH 549
QY 635 SINGVYFDSL-OLSYCLHEVAWYLLISGAQDFLSVFESGTEFKHKKYVEDTLTLPFS 693
Db 550 AINGFLFSNLPRLDMCKKGDYAMHLLGLTETDVGAVFQGTQVLOQMRGAANLFPHT 609
QY 694 GETVMSMENPGLWILGCHNSDFNRNGMTALIKVSSCDKNTGDIYEDSYEDISAYLSKN 753
Db 610 FYWAIQPDNLGTFLIYCOAGSHRAGRAIYVNSQC----- 647
QY 754 NAIERFSQGNPVLKRRQREITRTLLQSDQEEIDYDDITISVEMKKEDFDIYDEDENSP 813
Db 647 ----PGHQ-----ATP 653
QY 814 RSFOKTRHYFIAAVERLMDYGMSSS-PHYLRNRAOSGSV-----POFKRVY 859
Db 654 RQRYQARITYYIAAEVEVMDYCPDRSWEHMHNOSEKDSYGIIFLSNKDGLGSRKKAY 713
QY 860 FQEFIDGSGFTQPLXRGELNHLILGPYIAEVEDNINMVFERNQASRPYSFS 912
Db 714 FREYTDGTRIPRPRGPEHILGILPILKEVGDITLVFKNASRPYSVAH 766

RESULT 8
014286 PRELIMINARY; PRT; 216 AA.
AC 014286;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DE 01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE COAGULATION FACTOR VIII ASSOCIATED PROTEIN B.
GN F8B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE; 93052386.
RX LEVINSON B., KENNRICK S., GAMEL P., FISHER K., GITSCHER J.;
RT "Evidence for a third transcript from the human factor VIII gene.";
Genomics 14:585-589(1992).

```


DR EMBL; M90707; AAA58466.1; -;
DR PFAM; PF00754; F5_F8_type_C; 1.
SQ SEQUENCE 216 AA: 24641 MW; BBBCB9A5 CRC32;

Query Match	14.28;	Score 1104;	DB 4;	Length 216;
Best Local Similarity	100.0%;	Pred. No. 2e-76;		
Matches 208; Conservative	0;	Mismatches	0;	Gaps 0;

QY	1250	VFNNDSSGKKNHIFNPPILAIYIRLHPHYSIRSTLMLMELMGCDLNSCNPLOMESKA	1309
Db	9	VFNNDSSGKKNHIFNPPILAIYIRLHPHYSIRSTLMLMELMGCDLNSCNPLOMESKA	68
QY	1310	ISDQIATASSYFTNMFATNTPSKARLHLOGRSNAAMPVNNPREMLQVDFQTKMKTGYT	1369
Db	69	ISDQIATASSYFTNMFATNTPSKARLHLOGRSNAAMPVNNPREMLQVDFQTKMKTGYT	128
QY	1370	TQGVKSLLTSMYKEFLISSODGHQWTFEFGNGKVKVEFGQNDSTFPVNSLDPPLTLR	1429
Db	129	TQGVKSLLTSMYKEFLISSODGHQWTFEFGNGKVKVEFGQNDSTFPVNSLDPPLTLR	188
QY	1430	YLRHPQSWYHQALRMEVLGCENADLY	1457
Db	189	YLRHPQSWYHQALRMEVLGCENADLY	216

```

01JUN91      9
ID          043854      PRELIMINARY;      PRT;      480 AA.
AC          043854;
DT          01-JUN-1998 (TEMBLrel. 06, Created)
DT          01-JUN-1998 (TEMBLrel. 06, Last sequence update)
DT          01-MAY-1999 (TEMBLrel. 10, Last annotation update)
DE          INEGRIN BINDING PROTEIN DEL-1.
GN          DEL1.
OS          Homo sapiens (Human) .
OC          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC          Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN          [1]
RP          SEQUENCE FROM N.A.
RC          TISSUE=LUNG.
RX          MEDLINE; 98083109.
RA          HIDAI C., ZUPANCIC T.J., PENTA K., MIKHAIL A., KAWANA M.,
RA          OUBERKHOUS E.E., AOKA Y., FUKAGAWA M., MATSUI Y., PLATIKRA D.],
RA          AUDREBACH R., HOGAN B.L.M., SNODGRASS R., QUERTENMOUS T.;
RT          "Cloning and characterization of developmental endothelial locus-1:
RT          an embryonic endothelial cell protein that binds the alphavbeta3
RT          integrin receptor.";
RT          Genes Dev. 12:21-33(1998).
RL          EMBL; U70312; AAC02648.1; -.
RF          PFAM; PFO0008; EGF; 3.
RF          PFAM; PFO0754; F5_F8_type.C; 2.
DR          PROSITE; PS01187; EGF_CA_1.
DR          PROSITE; PS01285; FASEC_1.
DR          PROSITE; PS01286; FASEC_2.
DR          Integrin; Glycoprotein; EGF-like domain.
SQ          SEQUENCE 480 AA; 53765 MW; B59BFI3E CRC32;

```

	Query Match	8.5%	Score 662.5;	DB 4;	Length 480;
	Best Local Similarity	44.6%;	Pred. No. 3e-42;		
	Matches 145; Conservative	47;	Mismatches 118;	Indels 15;	Gaps 5
OY	1145 KCQTVLGMASGHIRDFQITASGOY-----GWMAPKLRLHYSGSINAWSTE--PFSNI	1196			
	: : :	: :	:		
Dd	157 KCSGLGIGGIINSNOQTIAASTTHRALFQLGWYPYATLNKKGLNMTAENADWPMPI	216			
OY	1197 KYDLAPMIIHSIKTQGAKOFKSSLYISOFIITSLDGRKWOTYRGNSTGLTWFGVND	1256			
	:::: : : : : :	: :	:		
Dd	217 QNLQRKMHVLTGVITGGARICSPETIKRYKLAYSNDGTWAMYKKYGINDMEVFEGNID	276			
OY	1257 SSGIKHNIFNPILIARYIRLAPTHYSIRSTLMELMGCDLNSCMLDGESKAISDAQIT	1316			
	:: : : :: :	:	:		

Db	277	NNTPRANSFTPIKAQYVRLYPQVCRHCHLRMELLCGCELSGCSSEPLGKSGHIDQYIT	336
QY	1317	ASSYE----TWMPATWPSBKRLHLOGRSNAMPVYNNPKEMLOYDQFQKTMGVGTGTTG	1376
Db	337	ASSITRLNMDMF-TWEPKRAKLEKQKVNAMTSGHNDOSQMLQYDULLPYTKYGLITGG	395
QY	1373	VKSLTSTMYKREFFLISSBODQHWTLL--FONGKVKYVQGNODSFTPVVNSLDPRLITRY	1430
Db	396	AKDFGHVQFVGSYKLAVNSDGEHNTVYQDEKQRDKDYQGNFNFDNHRKKNVDPRLYAH	455
QY	1431	LRIHPQSVWFOALRMELYGCEAOD	1455
Db	456	IRILPWSYWGRIITRSELGCTEE	480

```

RESULT 10
035474
ID 035474; PRELIMINARY; PRT; 480 AA.
AC 035474;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE DELI MAJOR SPICE VARIANT.
GN DELI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA QUENTERROUS T., HIDAI C., ZUPANCIC T., PENTA K., MIKAIL A.,
RA KAWANA M., QUENTERROUS E.E., AOKA Y., FUKUKAWA M., MATSUI Y.,
RA PLATKA D., AUERBACH R., HOGAN B.L.M., SNOGRASS R.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF031524; AAB86585.1; -
DR PFAM: PF00008; EGF_3.
DR PFAM: PF00754; F5_F8_type_C_2.
DR PROSITE: PS01187; EGF_CA_1.
DR PROSITE: PS01285; FASEC_1; 2.
DR PROSITE: PS01286; FASEC_2; 2.
DR Glycoprotein; EGF-like domain.
SQ SEQUENCE 480 AA: 53740 MW: 3286876 CRC32:

```

Query Match	8.4%	Score 655.5	DB 11	Length 480
Best Local Similarity	44.6%	Pred. No. 1e-41		
Matches 145	Conservative 45	Mismatches 120	Indels 15	Gaps 5
QY	1145	KQOTPLGMSASHIRDPQTASQY-----GGMARKLRLHSGSINMSTYE--PFSWI	1196	
		: : : : : : : : :		
Db	157	KOSGPGLIEGGILISMOQITLASSTHNALFGLRWYPRYLRLNKGILIMTALENDRWPI	216	
QY	1197	KYDLAPLRIHIGIKNGAROKFSSLYISOFLIMTSLDGKMQOTYRGNSGTGLMEFFGND	1256	
		::: : : : : : :		
Db	217	QINLQKKMAYTVITOGAKRIGSPETIYSIKATAYSNDGKTMAIVYKGTINEMVRGND	276	
QY	1257	SSGIKHNIFNPPIIARYIRLPHTRYISIRSLTLMELMGCDLSCSNPLGMSKALSDAQIT	1316	
		::: : : : :		
Db	277	NMTPRANSTPPIKQYVALYPLQICRHRHOTLMELLGELSCSPRLGKSKCHIDYQIT	336	
QY	1317	ASSYF---TNMFATWSPSKARLHLQGRSNARPOVNNPKEMLOYDFQTKMVTGVTQG	1372	
		: : : :		
Db	337	ASSVFETLLNMF-TWEPKARLDKQKGNAMVTSCHNDQSQWLYDLVPRTVGTITQG	395	
QY	1373	VNSLTSMTYKEFFLISSSDQHQWTLF--FQNGKAYKIVQGNODSTPRVYNSLDPLRLRY	1430	
		: : : : : : : : : : : : : : :		
Db	396	AADFQHVQDVGYSYKAYSNDGEHMMVHODEKORKDKVFOGNEFDNTIHRKNVIDPPIYARF	455	
QY	1431	LRIHPQSWHOTALMEYVLGCAQD	1455	
		:: : : : :		
Db	456	IRILPMSWIGRTLISELGAEE	480	

077718 ID 077718 PRELIMINARY; PRT; 363 AA.
AC 060462;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DE SPERM-MEMBRANE ASSOCIATED PROTEIN P47 (FRAGMENT).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Perissodactyla; Equidae; Equus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RC GENITEL M., TOEPFER-PETERSEN E.;
RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF010121; GAO09010.1; -
DR PFAM: PF00008; EGF: 1.
DR PFAM: PF00754; F5_F8_Type_C; 2.
Sperm.
FT NON_TER 1 1
FT NON_TER 363 363
SQ SEQUENCE 363 AA; 40744 MW; B470B46A CRC32;
Query Match 6.6%; Score 514.5; DB 6; Length 363;
Best Local Similarity 36.0%; Pred. No. 3.5e-31;
Matches 126; Conservative 49; Mismatches 118; Indels 57; Gaps 11;
QY 1124 ECLIGELHMGSLFLVY-----SNKOTPLGMSGHRDROITASGOY-- 1169
DB 42 ECHVIDDSHG--DVEYOYCSCPRGTGTHCETTCAMPJGEMGATADQISMSYFYG 99
QY 1169 -----GOWAPKLARLHSGSINAWST---KEPSWIKVDLAPMILHIGIKTGAROKFS 1220
DB 100 FMGLGRVPELRLHRTGIYANWATASNYDKNP--WIOVNLRRKRYGVYTGASRGSTA 157
QY 1221 LYSQFLIMSLDGKKWQYTRGNSGTGLMYPFGVNDSSGKHNITNPPIARIQLHPTH 1280
DB 158 EYKTFEYVASVGRKQFTRDAGDSKDKYGVNDSGLKVMFVPLEVTVRLVP-- 216
QY 1281 YSIRSTARMELMGCDLNSGMSPLGMEKASIDAOITASSYF---TMMFATWSPSKRLH 1336
DB 216 -----LACH-HGCTRP-----DROITASSYTRTGLNAF-SWIFRYRLD 253
QY 1337 LOGRSNARPOVNNPKEMLOYDFQTKMYTGVTTQGYKSLTSMYKEFLISSODGHOV 1396
DB 254 KOGFNNMTAQSNSASEMLDDDLGSKQEVGTITQGARDCGHQYDAKYVSHNDGANW 313
QY 1397 TLFFQN--GVKYVQGNODSTFPVNSLDPPLTRITRIHPQSWHOIAL 1444
DB 314 TEYRDORADSKITFLGNLDNNSHKMMEETPEFLARFVRLIFVAMHRRITL 363
RESULT 12
060462 ID 060462 PRELIMINARY; PRT; 931 AA.
AC 060462;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DE 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE VASCULAR ENDOTHELIAL CELL GROWTH FACTOR 165 RECEPTOR 2.
GN VEGF165R2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BREAST;
RC MEDLINE: 9818099.
RA SOKER S., TAKASHIMA S., MIAO H.Q., NEUFELD G., KLAGSBRUN M.;
RT "Neuropilin-1 is expressed by endothelial and tumor cells as an
RT isoform-specific receptor for vascular endothelial growth factor.";
RL Cell 92:735-745(1998).

DR EMBL: AF016096; AAC12922.1; -
DR PFAM: PF00431; CUB; 2.
DR PFAM: PF00754; F5_F8_Type_C; 2.
DR PFAM: PF00629; NAM; 1.
SQ SEQUENCE 931 AA; 104830 MW; 912262F2 CRC32;
Query Match 6.0%; Score 469.5; DB 4; Length 931;
Best Local Similarity 28.5%; Pred. No. 4.4e-27;
Matches 166; Conservative 87; Mismatches 185; Indels 145; Gaps 32;
QY 953 DEPDCK-AMATSYDVLK-----DVHSLGILPLVCHNTL-----NNAHQGYVVOEFA 1002
DB 79 EKHDCYDFIEIRDDSEADLLGRKCGNIAPPTIISGSMYIKFTSDYAROGA--GFS 136
QY 1003 LEFTIPDEFKSWFTENMRNCRAPCNIOEDPTFKENYRFAINGYIMDTPLGLVMAOD 1062
DB 137 LKREIF-KTGS-----EDCKNFTSP-NGTIESPGEKYP----- 171
QY 1063 QRLRWYLLSMGSNENIHSIFSGHVFYVARKKEEYKALYNLPGVFETVEMLPKAG-- 1120
DB 171 -----HNDCFTFLIAPKMEIILQFL--IFD-LEHDPLOVGEED 207
QY 1120 -----IW-----RYECLIGEH-----LHAGMST-----LFLV 1141
DB 208 CRYDMDLMDGIPHVGPIIGKCYGKTPELSRSTGILSTFTMDAAVAKDGFSAITYLV 267
QY 1142 YSN-----KQTPPLGMSGHRDFOITASGOY--GOWAPKLARLHSGSINAW-----ST 1189
DB 268 HOEPLNFCQNNPPLGMSGRINAEQISASSTYSDGKWTPQSRHL--GDNDMTNLDJN 325
QY 1190 KEPSWIKVDLAPMILHIGIKTGA--ROKESLSYISQFTIMYSLDGKKWQYTRGNSGT 1247
DB 326 KE--YLOVDFELTMTLTATOGAISRETONGYKYSKLEVSTNGEDMWNVYRHGKNH 382
QY 1248 LMVFGVNDSSGKHNITNPPIARIQLHPTHYSIRSTARMELMGCDLNS--CSMPLGM 1305
DB 382 -KVFOANNDAATEVYLLKLPALITRFVRLPOTWHSGLIRLELFCRYTADPCSMMLGM 440
QY 1306 ESKAISDAOITASSYTNMFATWSPSKARLHLOGRSNAN--RPOVNNPKEMLOYDFORT 1362
DB 441 LSGLDLSQISASS--TQGY-LMSPEARL-VSSNR-GMFPRIPQAPPEMLLOYDLGTP 495
QY 1363 MKYGVTTG-----VKSLSMAYKEFLISSODGHOVTLFPONGKY--KVFOGND 1413
DB 496 KTYKGVYIIGARGGDSITVEARAFYKRFKYSYSLGKDMW-YIDPRTQOQRLFEGBNH 554
QY 1414 SEPVPVNSLDPPLTRITRIHPQSWHO--IALRMEVLAGCAD 1455
DB 555 YDIPDIRFD-PIPAQYVRYFERKMSPAIGMKRLVGLGCDWTD 596
RESULT 13
014820 ID 014820 PRELIMINARY; PRT; 909 AA.
AC 014820;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DE 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE NEUROPILIN-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 9747088.
RA CHEN H., CHEDORAL A., HE Z.-G., GOODMAN C.S., TESSIER-LAVIGNE M.;
RT "Neuropilin-2, a novel member of the neuropilin family, is a high
RT affinity receptor for the semaphorins Sema E and Sema IV but not Sema
RT III.";
RL Neuron 19:547-559(1997).
DR EMBL: AF022859; AAC51788.1; -
DR PFAM: PF00431; CUB; 2.

DR PFAM: PF00754; F5_F8_type_C; 2.
DR PFAM: PF00629; MAM; 1.
SQ SEQUENCE 909 AA; 102213 MW; A5F7D0CD CRC32;

Query Match 6.0%; Score 469.5; DB 4; Length 909;
Best Local Similarity 28.5%; Pred. No. 4,2e-27;
Matches 166; Conservative 87; Mismatches 185; Indels 145; Gaps 32;

QY 953 DEFCK-AMAFSDVLEK---DVHSGLIGPLLYCHTNTL-----NPAIGROYTVOEFA 1002
DB 79 EKHCKKDFEIRGDSESDADLGRHCGNINPPTIISGSMXYIKFTSDYAROGA--GFS 136
QY 1003 LEFTIPDETSMYFTENMERCAPCNQIOMEDPFKEKRYRHAINGYIMTLPLVMAOD 1062
DB 137 LRYEIF-KTGS-----ECSKFTSP-NGTIESGFPPEKYP----- 171
QY 1063 QRIRWYLLSMGSMENIHSHFSGHVFYRKKEEYKMAIYLPVGFETVEMLPKAG-- 1120
DB 171 -----NHLDCTFTILAKPKMEIILQFL---IFD-LEHDPLOVGECD 207
QY 1120 -----IW-----RVECLIGEH-----LHAGMST-----LFLV 1141
DB 208 CKYDMLDIMGIPHYGPIGKYGCTKPSLRSSTGILSLFHTDMAVAKDGFSAARYLV 267
QY 1142 YSN-----KOOTPLGMASGHIHDFQITASGOY--GOMAPKLARLHSGSINAM-----ST 1189
DB 268 HOEPLNFQCNVPLGMSGRANEOISASTYSIDGRWTPPOOSRHH--GDDNGWTPNIDSN 325
QY 1190 KEPSWIKVDLAPMIHIGIKTGA--ROKFSLYISOFTIMYSLDGKKQOTYRGNSTGT 1247
DB 326 KE---YLOVDLRFMLTALATOGAISRENGYVASYLKEVSTNEDDMVYHGGKNH- 382
QY 1248 LMVFGVNDSSGIRHNIFNPPIIARYIRLPHTHYSINSTLMELMGCDLNS--CSMPLGM 1305
DB 382 -KYQANNDAIEVNLKHLAPLRLFRVIRIPQTHSGIALRLLEFGCRVTDAPCSNMLGM 440
QY 1306 ESKAISDAOTIASSYFTNMATSPSKARLHLOGRSNAM--RQVNNPKMLOVDPOKT 1362
DB 441 LSGLIASQISASS--TOEY-LMSPSARL-VSSRS-GWPRIPPOAOGEEMLVODVGT 495
QY 1363 MKYGVTTQG-----VKSLLTSMYKREPLISSQDGHQWTLFFONGKV--KYFOGNO 1413
DB 496 KTVGAVIIQARGDSTIAVEARAFVRKFKYSYSLNGKDE-YIQDPRTQPKLFEGBMH 554
QY 1414 SFTYVNSLDPPLRLRLRHPOSWHQ-IALRNEVLGCEAD 1455
DB 555 YDPPDIRFD-PIPAQYRVYPERMSPAGIGMRLEVIGCDWTD 596

RESULT 14
Q14821
ID 014821 PRELIMINARY; PRT; 926 AA.
AC 014821
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE NEUROFILIN-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RX MEDLINE; 97470888.
RA KOLODKIN A.L., LEVENGOOD D.V., ROWE E.G., TAI Y.-T., GIGER R.J.,
RT "Neurofilin-2, a novel member of the neurofilin family, is a high
affinity receptor for the semaphorins Sema E and Sema IV but not Sema
III.".
RL Neuron 10:547-559(1997).
DR EMBL; AF022860; AACSI1789.1; -.
DR PFAM; PF00431; CUB; 2.
DR PFAM; PF00754; F5_F8_type_C; 2.
DR PFAM; PF00629; MAM; 1.

SQ SEQUENCE 926 AA; 104253 MW; 4C812073 CRC32;

Query Match 6.0%; Score 469.5; DB 4; Length 926;
Best Local Similarity 28.5%; Pred. No. 4,3e-27;
Matches 166; Conservative 87; Mismatches 185; Indels 145; Gaps 32;

QY 953 DEFCK-AMAFSDVLEK---DVHSGLIGPLLYCHTNTL-----NPAIGROYTVOEFA 1002
DB 79 EKHCKKDFEIRGDSESDADLGRHCGNINPPTIISGSMXYIKFTSDYAROGA--GFS 136
QY 1003 LEFTIPDETSMYFTENMERCAPCNQIOMEDPFKEKRYRHAINGYIMTLPLVMAOD 1062
DB 137 LRYEIF-KTGS-----ECSKFTSP-NGTIESGFPPEKYP----- 171
QY 1063 QRIRWYLLSMGSMENIHSHFSGHVFYRKKEEYKMAIYLPVGFETVEMLPKAG-- 1120
DB 171 -----NHLDCTFTILAKPKMEIILQFL---IFD-LEHDPLOVGECD 207
QY 1120 -----IW-----RVECLIGEH-----LHAGMST-----LFLV 1141
DB 208 CKYDMLDIMGIPHYGPIGKYGCTKPSLRSSTGILSLFHTDMAVAKDGFSAARYLV 267
QY 1142 YSN-----KOOTPLGMASGHIHDFQITASGOY--GOMAPKLARLHSGSINAM-----ST 1189
DB 268 HOEPLNFQCNVPLGMSGRANEOISASTYSIDGRWTPPOOSRHH--GDDNGWTPNIDSN 325
QY 1190 KEPSWIKVDLAPMIHIGIKTGA--ROKFSLYISOFTIMYSLDGKKQOTYRGNSTGT 1247
DB 326 KE---YLOVDLRFMLTALATOGAISRENGYVASYLKEVSTNEDDMVYHGGKNH- 382
QY 1248 LMVFGVNDSSGIRHNIFNPPIIARYIRLPHTHYSINSTLMELMGCDLNS--CSMPLGM 1305
DB 382 -KYQANNDAIEVNLKHLAPLRLFRVIRIPQTHSGIALRLLEFGCRVTDAPCSNMLGM 440
QY 1306 ESKAISDAOTIASSYFTNMATSPSKARLHLOGRSNAM--RQVNNPKMLOVDPOKT 1362
DB 441 LSGLIASQISASS--TOEY-LMSPSARL-VSSRS-GWPRIPPOAOGEEMLVODVGT 495
QY 1363 MKYGVTTQG-----VKSLLTSMYKREPLISSQDGHQWTLFFONGKV--KYFOGNO 1413
DB 496 KTVGAVIIQARGDSTIAVEARAFVRKFKYSYSLNGKDE-YIQDPRTQPKLFEGBMH 554
QY 1414 SFTYVNSLDPPLRLRLRHPOSWHQ-IALRNEVLGCEAD 1455
DB 555 YDPPDIRFD-PIPAQYRVYPERMSPAGIGMRLEVIGCDWTD 596

RESULT 15
O35276
ID 035276 PRELIMINARY; PRT; 925 AA.
AC 035276
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE NEUROFILIN-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RX MEDLINE; 97433085.
RA KOLODKIN A.L., LEVENGOOD D.V., ROWE E.G., TAI Y.-T., GIGER R.J.,
RT "Neurofilin is a semaphorin III receptor.".
RL Cell 90:753-762(1997).
DR EMBL; AF016297; AAC53338.1; -.
DR PFAM; PF00431; CUB; 2.
DR PFAM; PF00754; F5_F8_type_C; 2.
DR PFAM; PF00629; MAM; 1.
SQ SEQUENCE 925 AA; 103896 MW; B097CCCA CRC32;

Query Match 6.0%; Score 464.5; DB 11; Length 925;
Best Local Similarity 28.6%; Pred. No. 1e-26;
Matches 167; Conservative 85; Mismatches 186; Indels 145; Gaps 33;

QY 953 DEFDCK-AMAYFSDVDEK---DVHSGLLGPLLVCHTNTL-----NPAHQRYTVQGEFA 1002
DB 79 EKHCKYDFEIRGDSEADLLGKHCGNIAPTLISSGSVLYIKFTSDYARQGA--GFS 136
QY 1003 LEFTIDEFTKSWYFTEMENRCAPCNIOEDPTFKENYFFHAINGYIMDTLPGLVMAQD 1062
DB 137 LRYEIF-KTGS---EDCSKNFTSP-NGTIESPGFPEKYP----- 171
QY 1063 QRIRWYLLSGNSMENIHSHFSGHVFTVRKKEEYKMALYNLPGVFETVEMLPKAG-- 1120
DB 171 -----HNLDCTFTILAKPRMEILQFL--TFD-LEHDPLOVGECD 207
1120 -----IW-----RVECLIGEH-----LAGMST-----LFLV 1141
208 CKYDMLDIDWGIPIVHGPIGKYGCTKTPSKLRSSGTGLSLTFHTDMAVARDGFSARYLV 267
QY 1142 YSN-----KCOTPLGMASGHRDFOITASGOY--GOWAPKLARLHYSGSINAW-----ST 1189
DB 268 HQEPPEPFCNAPLGMESGRIANEQIASSTFSDGRNTPOOSRLH--GDDNGWTPNVDN 325
QY 1190 KEPEFWIKVDLLAPMIHGIRKTOGA--RQKFSLLYISOFLIMYSLDGKKWQTYRGNSTGT 1247
DB 326 KE---YLQVDRFLTMLTALATOGAISRETOQKGYVKSYKLEVSTNGEDMWVYRHGKNH- 382
QY 1248 LMVPEFGVNDSSGICKHNFNPPIIARLRLHPTHSIRSITLAMELMGCDLNS--CSMPLGM 1305
DB 382 -KVFQANNDADELVLNKLHPLRLFRIRPOTWHLGIALRLLELFGCRVTDAPCSNMLGM 440
QY 1306 ESKAISDAQITASSYFTNMFTWSPSKARLHLOGRSNAW--RPOVNNPKEMLOVDFOKT 1362
DB 441 LSGLIADTOISASS--TREY-LWSPSARL-VSSRS--GMFPRNPQAQPGCEWLDVLTGP 495
QY 1363 MKVTGVTTOGVK--SLITSM---YKEFLISSQDGHQWTLFPQNGKV--KVFQGNOD 1413
DB 496 KTVKGVYIIOGARGDSTIAMERAFVRRFKVYSYSLNGDME-YIODPRTQOPKLEEGNMH 554
QY 1414 SFTPVVNSLDPELLTRILIRIHPOSMVHQ-IALRMEVLGCEAOD 1455
DB 555 YDTPDIRFE-PVPAQYRVYPERMSPAGIGMRLEVLGCDWTD 596

rch completed: August 19, 1999, 12:40:39
Job time: 2334 sec


```

Db 241 AASARAPKMTVNGVYNRSLPGLIGCHRSVYWHVIGMGTTPVHSLFLEGHTFLVRNH 300
QY 301 ROASLEISPIITELTAQTLMDLMDLGOFLFCHLISHQHDGMEAYVYKVDCEPEEQLRMKNNE 360
Db 301 ROASLEISPIITELTAQTLMDLMDLGOFLFCHLISHQHDGMEAYVYKVDCEPEEQLRMKNNE 360
QY 361 EAEYDDDLTSEMDEVYFEDDNDSPSFIOIRSVAKKHPTKVWHYIAAEEDMDAPLVIA 420
Db 361 EAEYDDDLTSEMDEVYFEDDNDSPSFIOIRSVAKKHPTKVWHYIAAEEDMDAPLVIA 420
QY 421 PDDRSYKQYLNGPQIRIGRKYKKVRFMAVYDETFKIREALIOHESGLIGPLLYGEVDTL 480
Db 421 PDDRSYKQYLNGPQIRIGRKYKKVRFMAVYDETFKIREALIOHESGLIGPLLYGEVDTL 480
QY 481 LIIFKNQASRPYNIYPHGTTDVRPLYSRRLPKGVNHLKDFLPELIRKYMTVYVEDGP 540
Db 481 LIIFKNQASRPYNIYPHGTTDVRPLYSRRLPKGVNHLKDFLPELIRKYMTVYVEDGP 540
QY 541 TKSDRCLTRYSSVFNMERDLASGLIGPLLYCYKESVDQKNOJMSDKRNVILFVDE 600
Db 541 TKSDRCLTRYSSVFNMERDLASGLIGPLLYCYKESVDQKNOJMSDKRNVILFVDE 600
QY 601 NRSWYLTENIQRFLEPNPAGVLEDEPEFQASNIMHSINGYVEDSLQSVCLHEVAIWTLS 660
Db 601 NRSWYLTENIQRFLEPNPAGVLEDEPEFQASNIMHSINGYVEDSLQSVCLHEVAIWTLS 660
QY 661 IGAQDFELSVFSGYTFKRMKYEDTLTFPFSGEFVPMENPGLMILGCHNSDFRNRG 720
Db 661 IGAQDFELSVFSGYTFKRMKYEDTLTFPFSGEFVPMENPGLMILGCHNSDFRNRG 720
QY 721 MTALKVSSCDKNTGDYEDSYEDISAYLLSKNNAIEPRSFQNPVYLKRHOEITRTTL 780
Db 721 MTALKVSSCDKNTGDYEDSYEDISAYLLSKNNAIEPRSFQNPVYLKRHOEITRTTL 780
QY 781 QSDQEEIDYDITISVEMKEDFDIYDEDNOSPRFOKTRHYFLAVERLMDYGSSSP 840
Db 781 QSDQEEIDYDITISVEMKEDFDIYDEDNOSPRFOKTRHYFLAVERLMDYGSSSP 840
QY 841 HYLNRNAGSGVPOPKVYVOEFTDGSFTOPLYRGELNHLGLGPIYAEVEDNIMVTF 900
Db 841 HYLNRNAGSGVPOPKVYVOEFTDGSFTOPLYRGELNHLGLGPIYAEVEDNIMVTF 900
QY 901 RNQASRPYSFSSLSIYEEDQKAPRKNFYKPNETKYFMVQOHMAPTDEDFCKAW 960
Db 901 RNQASRPYSFSSLSIYEEDQKAPRKNFYKPNETKYFMVQOHMAPTDEDFCKAW 960
QY 961 AYSVDVLEKDVHSLIGPLLYCHTNTLPAHGRQVTOEFALFETIPEETKSWYTEMA 1020
Db 961 AYSVDVLEKDVHSLIGPLLYCHTNTLPAHGRQVTOEFALFETIPEETKSWYTEMA 1020
QY 1021 ERNCRAPCNIOEMDPTEFKENYRFHAINGYIMDTLPGIVNAODORIRMYLLSMGSNENIHS 1080
Db 1021 ERNCRAPCNIOEMDPTEFKENYRFHAINGYIMDTLPGIVNAODORIRMYLLSMGSNENIHS 1080
QY 1081 IHESGVVYFRKKEEKMAALYNYPGVFEVEMLPKAGIWRRECLIGELHAGMSTLFL 1140
Db 1081 IHESGVVYFRKKEEKMAALYNYPGVFEVEMLPKAGIWRRECLIGELHAGMSTLFL 1140
QY 1141 VYSNKCOTPLGMAAGHIRDFOITASGOYGOMAPKLARLHSGSINAMSTKEPSSWIKVDL 1200
Db 1141 VYSNKCOTPLGMAAGHIRDFOITASGOYGOMAPKLARLHSGSINAMSTKEPSSWIKVDL 1200
QY 1201 LAMMIHIGITQAGAROKFSSLSYISQITIMYSLDGKMQYRGNSTGTLMVFGNVDSSGI 1260
Db 1201 LAMMIHIGITQAGAROKFSSLSYISQITIMYSLDGKMQYRGNSTGTLMVFGNVDSSGI 1260
QY 1261 KAHINPPIIARIYRIHPHYSTRJLRMELMGCDLNSCAMPJGMEKASDQOIRASSY 1320
Db 1261 KAHINPPIIARIYRIHPHYSTRJLRMELMGCDLNSCAMPJGMEKASDQOIRASSY 1320
QY 1321 FTJNMFATWSPSKARLHLQGSNMAWRPQVNNPKEMLQVDFQRTKMYGVTTQGVKSLTSM 1380
Db 1321 FTJNMFATWSPSKARLHLQGSNMAWRPQVNNPKEMLQVDFQRTKMYGVTTQGVKSLTSM 1380

```

```

QY 1381 YKEFLISSQDCHQWTLFFQNGKVKVFQGNODSFYVNSLDPPLTRTYLRIRHQSWMH 1440
Db 1381 YKEFLISSQDCHQWTLFFQNGKVKVFQGNODSFYVNSLDPPLTRTYLRIRHQSWMH 1440
QY 1441 QIALRMEVLGCEAODLY 1457
Db 1441 QIALRMEVLGCEAODLY 1457

RESULT 2
W46246
ID W46246 standard; Protein; 1457 AA.
AC W46246;
AD 06-AUG-1998 (first entry)
DE Human factor VIII beta-domain deleted SQN deletion protein sequence.
KW Replication defective; recombinant retrovirus; RRV; therapeutic protein;
KW haemophilia; thrombosis; hypercoagulable disorder; liver disease; human;
KW cystic fibrosis; Duchenne's muscular dystrophy; hypercholesterolemia;
KW hypopituitarism; adenine deaminase deficiency; HIV infection; anaemia;
KW Guacher's syndrome; high blood pressure; Alzheimer's disease; autoimmune;
KW inflammatory disease; factor VIII.
OS Homo sapiens.
PN MO9800541-A2.
PD 08-JAN-1998.
PE 02-JUL-1997; U11784.
PR 04-JUN-1997; US-869309.
PR 03-JUL-1996; US-645601.
PR 13-AUG-1996; US-696381.
PA (CHIR ) CHIRON CORP.
PI Allen JR, Barber JR, Boder M, Chang SM, Chong K,
PI De la Vega D, Depoloni, Greengard J, Hsu DC, Ibanez CE,
PI Jolly DJ, Lee R, Mittelstaedt DM, Prussak CE, Respass JG,
DR WPI: 98-086966/08.
DR N-PSDB: V19581.
PT New replication defective recombinant retro-viruses - which can be
PT administered to provide long term systemic expression of therapeutic
PT protein in blood, useful in, e.g. treating hyper-coagulable
PT disorders
PS Example 28; Pages 213-217; 272pp; English.
SC This is the beta-domain deleted SQN deletion protein of human factor
SC VIII. The encoding DNA is used to construct recombinant retroviral
SC vectors expressing human factor VIII. The invention provides the
SC preparation of replication defective recombinant retrovirus (RRV)
SC degrading a therapeutic protein. The RRV preparation is resistant to
SC degradation by human complement and is capable of inducing long term
SC systemic expression of the therapeutic protein when administered
SC intravenously to a human. The long term systemic expression results in a
SC measurable level of the therapeutic protein being produced in the blood
SC of the human for a period of at least 30 days after the administration of
SC the RRV vector preparation. RRV's can be used for in vivo delivery of
SC therapeutic protein to treat, e.g. haemophilia A, haemophilia B,
SC thrombotic, hypercoagulable disorders, liver diseases such as hepatitis,
SC disorders such as thalassemia, phenylketonuria, Lesch-Nyhan syndrome,
SC severe combined immunodeficiency (SCID), cystic fibrosis, Duchenne's
SC muscular dystrophy, inherited emphysema, familial hypercholesterolemia,
SC diabetes, hypopituitarism, adenine deaminase deficiency, alpha1-
SC antitrypsin deficiency, Guacher's syndrome, anaemia, infections such as
SC HIV infection, high blood pressure, Alzheimer's disease, autoimmune or
SC inflammatory disease or graft versus host disease. RRV's are capable of
SC surviving inactivation in human serum thereby allowing efficient gene
SC transfer over prolonged periods of time.
SQ Sequence 1457 AA;

```

Query Match 100.0%; Score 7797; DB 1; Length 1457;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MOIEIETCFLLCLRPFCSATRRYILGAVELSWDMQSDLGEPLVDARFPVRPKSPFN 60
Db 1 MOIEIETCFLLCLRPFCSATRRYILGAVELSWDMQSDLGEPLVDARFPVRPKSPFN 60

```

```

QY 61 TSVYKKTLEFEEDHLEFNIAKRPBPMGLIPTIOAEVYDVTYTLKNMA SHPYSIAV 120
DB 61 TSVYKKTLEFEEDHLEFNIAKRPBPMGLIPTIOAEVYDVTYTLKNMA SHPYSIAV 120
QY 121 GVSYWKASGEAEYDDQTSQREKEDKVPFGSGHLYVMQVLKENGSMASDPLCTLYSH 180
DB 121 GVSYWKASGEAEYDDQTSQREKEDKVPFGSGHLYVMQVLKENGSMASDPLCTLYSH 180
QY 181 VDLVKDNLGSLIGALLVCRESSLAKEKTQTLHKFLLFAVDEGKSWHSETKNSLMODRD 240
DB 181 VDLVKDNLGSLIGALLVCRESSLAKEKTQTLHKFLLFAVDEGKSWHSETKNSLMODRD 240
QY 241 AASAPAMKMTVNGYVNRSLPGLGCHRSYVYHVIYMGSTPEVHSIFLBSHTLYVKNH 300
DB 241 AASAPAMKMTVNGYVNRSLPGLGCHRSYVYHVIYMGSTPEVHSIFLBSHTLYVKNH 300
QY 301 ROASLEISPIFLFQTLLMDLGOFLFCHTISHOHGMEAVKYVDSGPEEPOLMKKNE 360
DB 301 ROASLEISPIFLFQTLLMDLGOFLFCHTISHOHGMEAVKYVDSGPEEPOLMKKNE 360
QY 361 EAEDYDDLTDSEMDVAFRDDNSSFQIRSVAKKHRTWVHYIAEEDMDYAPLYLA 420
DB 361 EAEDYDDLTDSEMDVAFRDDNSSFQIRSVAKKHRTWVHYIAEEDMDYAPLYLA 420
QY 421 PDDRSYKSOYLNGPQRIGRKRYKATYDEFTKREAIQHEGSLIGPLLYGEVDTL 480
DB 421 PDDRSYKSOYLNGPQRIGRKRYKATYDEFTKREAIQHEGSLIGPLLYGEVDTL 480
QY 481 LIIFNQASRPNIYPHGTDVRLPYLSRRLPKGVKHLKDFPLPEIFRYKKTAVVEGDP 540
DB 481 LIIFNQASRPNIYPHGTDVRLPYLSRRLPKGVKHLKDFPLPEIFRYKKTAVVEGDP 540
QY 541 TKSDPRLTRYSSFVNMRDLASGLIPLLICYESVDORGNQIMSDKRNVIYLSVEDE 600
DB 541 TKSDPRLTRYSSFVNMRDLASGLIPLLICYESVDORGNQIMSDKRNVIYLSVEDE 600
QY 601 NRSWLTENIQRLPNPAGVQLEDEPEFQASNTMHSINGVPSLQSLCHVAVAYTIS 660
DB 601 NRSWLTENIQRLPNPAGVQLEDEPEFQASNTMHSINGVPSLQSLCHVAVAYTIS 660
QY 661 IGAQDFLSVFEFSGYTFKHKMYEDTLTFPSSGETVFESMNPGLMILGINSFRRNG 720
DB 661 IGAQDFLSVFEFSGYTFKHKMYEDTLTFPSSGETVFESMNPGLMILGINSFRRNG 720
QY 721 MTALLKSSCKNTGDIYEDSYEDISAVLLSKNNAIEPRFSQNPVLKRHRREITRTTL 780
DB 721 MTALLKSSCKNTGDIYEDSYEDISAVLLSKNNAIEPRFSQNPVLKRHRREITRTTL 780
QY 781 OSDQDEIDYDITISVEMKKEPDFDIYEDENOSPRSFQKTRHYFTAVERLMDYGMSSSP 840
DB 781 OSDQDEIDYDITISVEMKKEPDFDIYEDENOSPRSFQKTRHYFTAVERLMDYGMSSSP 840
QY 841 HYLNRRAAGSGVPOFKVYFOEFTGSGFTOPLYRGELENHGLLPYIRAEVEDIMYTF 900
DB 841 HYLNRRAAGSGVPOFKVYFOEFTGSGFTOPLYRGELENHGLLPYIRAEVEDIMYTF 900
QY 901 RNOASRPISFYSSLSIYEDDQROGAEPKRNKFNPKNETKTYFKKVOHMAPTDEDFCKAM 960
DB 901 RNOASRPISFYSSLSIYEDDQROGAEPKRNKFNPKNETKTYFKKVOHMAPTDEDFCKAM 960
QY 961 AYSFVVDLEKDVHSGLIGPLVCHTNTLPAHAGROYTVOEFALFTIEDETKSWYTEMN 1020
DB 961 AYSFVVDLEKDVHSGLIGPLVCHTNTLPAHAGROYTVOEFALFTIEDETKSWYTEMN 1020
QY 1021 EBNCRAPCIOMEDPTFKENYRFAHNGYIMDTLGLVMAOORIRKWTLLSGMSNHTS 1080
DB 1021 EBNCRAPCIOMEDPTFKENYRFAHNGYIMDTLGLVMAOORIRKWTLLSGMSNHTS 1080
QY 1081 IHFSGHVTVRKKEEYKMAALYMLPGVEFTEVEMLPKAGINRVECLIGEHLAGSTLFL 1140
DB 1081 IHFSGHVTVRKKEEYKMAALYMLPGVEFTEVEMLPKAGINRVECLIGEHLAGSTLFL 1140

```

```

QY 1141 VYSNKCOTPLGASGHTRPDQITASGOYGMAPKLARLHYSGSTINAMSTKEPSMIVDL 1200
DB 1141 VYSNKCOTPLGASGHTRPDQITASGOYGMAPKLARLHYSGSTINAMSTKEPSMIVDL 1200
QY 1201 LAPMIHGIKTQAGAROKRSSIYISQFIIMYSLDCKKMQTYGNSGTGLMVFEGVNDSSGI 1260
DB 1201 LAPMIHGIKTQAGAROKRSSIYISQFIIMYSLDCKKMQTYGNSGTGLMVFEGVNDSSGI 1260
QY 1261 KHNIFNPPIIARYIRLPHYHSIRSTLMELMGCDLNSCSMPLGMSKSAISDAQITASSY 1320
DB 1261 KHNIFNPPIIARYIRLPHYHSIRSTLMELMGCDLNSCSMPLGMSKSAISDAQITASSY 1320
QY 1321 FTNNFATWSPSKARLHLQGRBNAMPQVNNPKEMVLQVDFQRTKMTGYTGTGVSLLTSM 1380
DB 1321 FTNNFATWSPSKARLHLQGRBNAMPQVNNPKEMVLQVDFQRTKMTGYTGTGVSLLTSM 1380
QY 1381 YKFEFLISSODGHQWTLFPNGKVKYQGNDSFTPVNSLDPLRLRYIRIHQSMVH 1440
DB 1381 YKFEFLISSODGHQWTLFPNGKVKYQGNDSFTPVNSLDPLRLRYIRIHQSMVH 1440
QY 1441 QIALMEVYLGEADOLY 1457
DB 1441 QIALMEVYLGEADOLY 1457

RESULT 3
W23414
ID W23414 standard; Protein; 1471 AA.
AC W23414;
DE 08-APR-1998 (first entry)
DE Human B-domain deleted factor VIII protein.
KW Post-translational regulatory element; PRE: enhancer II; intronless gene;
KW surface antigen gene; cytoplasmic accumulation; targeted delivery;
KW near consensus splice sequence; blood coagulation factor; factor VIII;
KW factor IX.
OS Homo sapiens.
PN M0973394-A1.
PD 18-SEP-1997.
PR 10-MAR-1997; U03561.
PR 11-MAR-1996; U5-683839.
PA (IMMU-) IMMUNE RESPONSE CORP.
PI Bidlingmayer S, Ill CR:
DR WPI: 97-470874/43.
DR N-PSDB: 773164.
DR Vector for increased expression of intronless genes - comprises
PT intronless gene with at least one near consensus splice sequence, a
PT promoter and at least one viral cis-acting post-transcriptional
PT regulatory element
PS Example 1; Pages 31-36; 59pp; English.
CC The present sequence represents human B-domain deleted factor VIII
CC The cDNA encoding this protein also contains, 3' of the coding region, a
CC post-translational regulatory element (PRE) of the Hepatitis B virus.
CC PRE sequences have been shown to function in cis to increase the
CC steady-state levels of surface gene transcripts by facilitating
CC cytoplasmic accumulation of these transcripts. The above nucleic acid
CC sequence is part of a novel vector, comprising an intronless gene
CC containing 1 or more near consensus splice sequences operably linked to
CC a promoter sequence so that the gene is transcribed in a cell.
CC Intronless gene transcripts which contain near consensus splice site
CC sequences are believed to get tied up in the nucleus of the cell where
CC splicing occurs, rather than being transported to the cytoplasm where
CC they can be translated into proteins. The PRE sequences are transcribed
CC along with the gene, causing export of the gene transcript from the
CC nucleus into the cytoplasm of the cell. The vector can be used
CC to increase the expression of an intronless gene containing at least one
CC near consensus splice sites, preferably cDNA encoding a blood coagulation
CC factor, particularly factor VIII or IX. The complex allows the targeted
CC delivery of the vector to a specific cell, e.g. hepatocytes when the
CC 144nd is an asialoglycoprotein which binds the asialoglycoprotein
CC receptor present on their surface.
SQ Sequence 1471 AA:

```

Query Match 99.5%; Score 7759; DB 1; Length 1471;
 Best Local Similarity 98.8%; Pred. No. 0; Mismatches 3; Indels 14; Gaps 1;
 Matches 1454; Conservative 0;

1 MOEISTCFELLCIRFCFSATRRYYLGAVELSDWYQMSDGLGELPVDAREPPRVPKSPFN 60
 1 MYEELSTCFELLCIRFCFSATRRYYLGAVELSDWYQMSDGLGELPVDAREPPRVPKSPFN 60

61 TSVVYKTLFVEETHLLENIAKPRPMWGLLFTIOAEYDYVITLKNMASHPSLAHV 120
 61 TSVVYKTLFVEETHLLENIAKPRPMWGLLFTIOAEYDYVITLKNMASHPSLAHV 120

121 GSVYKASGAEDDQTSOREKEDKVPFGSGSHYMOVLKENGPAAPLCLITYSLH 180
 121 GSVYKASGAEDDQTSOREKEDKVPFGSGSHYMOVLKENGPAAPLCLITYSLH 180

181 VDLVNDLNSGLIGALLVYCREGSLAKERTQTLAKFILLFAVDEKSMHSETKNSLMQDRD 240
 181 VDLVNDLNSGLIGALLVYCREGSLAKERTQTLAKFILLFAVDEKSMHSETKNSLMQDRD 240

241 AASARAMPKHTVNGYVNSLPGLICGHRKSYWYIOMGTPEVHSIFLEGHTFLVRNH 300
 241 AASARAMPKHTVNGYVNSLPGLICGHRKSYWYIOMGTPEVHSIFLEGHTFLVRNH 300

301 ROASLEISPTFLTAOTLMDLGOFLFCHISSHODGMAAYKVPSCPEPOLRMKNNE 360
 301 ROASLEISPTFLTAOTLMDLGOFLFCHISSHODGMAAYKVPSCPEPOLRMKNNE 360

361 EADYDDDLTDSEMDVYRFDDNSPSFIQIRYAKKHPTWVHYIAAEEDMDYAPLVIA 420
 361 EADYDDDLTDSEMDVYRFDDNSPSFIQIRYAKKHPTWVHYIAAEEDMDYAPLVIA 420

421 PDRSXYQYTLNNGPQRIKRYKVRPMATDEFEKREAIQHSGLLGLLGEVDTL 480
 421 PDRSXYQYTLNNGPQRIKRYKVRPMATDEFEKREAIQHSGLLGLLGEVDTL 480

481 LIIFKNOASRPYNIYPGITDVRLYRRLPKGVKHLKDPILPGEIJKYKWTYVEDGP 540
 481 LIIFKNOASRPYNIYPGITDVRLYRRLPKGVKHLKDPILPGEIJKYKWTYVEDGP 540

541 TKSDPCLTRYSSFVNMERDLASGLIGPLLICKEVDKRGNOIMSDKNVILFSEVDE 600
 541 TKSDPCLTRYSSFVNMERDLASGLIGPLLICKEVDKRGNOIMSDKNVILFSEVDE 600

601 NRSWYLTENIORFLPNPAGVLEDPFOASNMHSINGYVDSIQLSVCLAEVAYWYILS 660
 601 NRSWYLTENIORFLPNPAGVLEDPFOASNMHSINGYVDSIQLSVCLAEVAYWYILS 660

661 IGAOTDFLSVFGSYTEFKHKWYEDTTLTFPFGSETFVMSMENGLMILGCHNSDFNRNG 720
 661 IGAOTDFLSVFGSYTEFKHKWYEDTTLTFPFGSETFVMSMENGLMILGCHNSDFNRNG 720

721 MTALKYSSCKKNGDYEDSYEDISALLSKNNAIEPRSSQNSRHPSTROKOFNATPP 780
 721 MTALKYSSCKKNGDYEDSYEDISALLSKNNAIEPRSSQNSRHPSTROKOFNATPP 780

767 VLKHHOREITRTTQSDOEIDYDITISVEAKKEDFDIYDEDEQSPRSFOKTRHYFIA 826
 767 VLKHHOREITRTTQSDOEIDYDITISVEAKKEDFDIYDEDEQSPRSFOKTRHYFIA 826

827 AVERLMDYGMSSPHVLNRAQSGVPOFKKVFQEOFTDGSFTOPLVINGELNEHLGLLGP 886
 827 AVERLMDYGMSSPHVLNRAQSGVPOFKKVFQEOFTDGSFTOPLVINGELNEHLGLLGP 886

887 YIRAEVENINIVFRNOSRPSYFSSLSIYEBDROGAERKRNKVNKNETKTFWAKQH 946
 887 YIRAEVENINIVFRNOSRPSYFSSLSIYEBDROGAERKRNKVNKNETKTFWAKQH 946

947 HMAPTKDEFDCKAMAFESADVLEKDVHSGLIGPLLVCHTNTLNPAHQVTVQEFALFT 1006
 947 HMAPTKDEFDCKAMAFESADVLEKDVHSGLIGPLLVCHTNTLNPAHQVTVQEFALFT 1006

1007 IFDETKSWYFTENNERCRAPCNIONMEDPTRENTREFHAINGYIMDTLPGLVMAODQIR 1066
 1007 IFDETKSWYFTENNERCRAPCNIONMEDPTRENTREFHAINGYIMDTLPGLVMAODQIR 1066

1021 IFDETKSWYFTENNERCRAPCNIONMEDPTRENTREFHAINGYIMDTLPGLVMAODQIR 1080
 1021 IFDETKSWYFTENNERCRAPCNIONMEDPTRENTREFHAINGYIMDTLPGLVMAODQIR 1080

1067 WYLLSMGSENIHSHFSGHVFYRKREKYMALXNYPGVEFEVEMLPKAGIWRRECL 1126
 1067 WYLLSMGSENIHSHFSGHVFYRKREKYMALXNYPGVEFEVEMLPKAGIWRRECL 1126

1081 WYLLSMGSENIHSHFSGHVFYRKREKYMALXNYPGVEFEVEMLPKAGIWRRECL 1140
 1081 WYLLSMGSENIHSHFSGHVFYRKREKYMALXNYPGVEFEVEMLPKAGIWRRECL 1140

1127 IGEHLHAGMSTFLVYSNKCOTPLGMAASHIRDFQITASGOYGOMAPKLRLAHYSGSINA 1186
 1127 IGEHLHAGMSTFLVYSNKCOTPLGMAASHIRDFQITASGOYGOMAPKLRLAHYSGSINA 1186

1141 IGEHLHAGMSTFLVYSNKCOTPLGMAASHIRDFQITASGOYGOMAPKLRLAHYSGSINA 1200
 1141 IGEHLHAGMSTFLVYSNKCOTPLGMAASHIRDFQITASGOYGOMAPKLRLAHYSGSINA 1200

1187 WSTKEPFSWIKVDLLAPMIHGIRTOGAROKSSLYTSQPTIMSUDGKKWQTYRGSTG 1246
 1187 WSTKEPFSWIKVDLLAPMIHGIRTOGAROKSSLYTSQPTIMSUDGKKWQTYRGSTG 1246

1201 WSTKEPFSWIKVDLLAPMIHGIRTOGAROKSSLYTSQPTIMSUDGKKWQTYRGSTG 1260
 1201 WSTKEPFSWIKVDLLAPMIHGIRTOGAROKSSLYTSQPTIMSUDGKKWQTYRGSTG 1260

1247 TLMVFGVNDSSGIRKHNFTNPITARYRLPPTYSIRSTLRMLMGDNLSCMPGME 1306
 1247 TLMVFGVNDSSGIRKHNFTNPITARYRLPPTYSIRSTLRMLMGDNLSCMPGME 1306

1261 TLMVFGVNDSSGIRKHNFTNPITARYRLPPTYSIRSTLRMLMGDNLSCMPGME 1320
 1261 TLMVFGVNDSSGIRKHNFTNPITARYRLPPTYSIRSTLRMLMGDNLSCMPGME 1320

1307 SKAISDAQITASSYFTNMFATWSPSKARLHLOGRSNMAPQVNNPKEMLQVDPQTKMYT 1366
 1307 SKAISDAQITASSYFTNMFATWSPSKARLHLOGRSNMAPQVNNPKEMLQVDPQTKMYT 1366

1321 SKAISDAQITASSYFTNMFATWSPSKARLHLOGRSNMAPQVNNPKEMLQVDPQTKMYT 1380
 1321 SKAISDAQITASSYFTNMFATWSPSKARLHLOGRSNMAPQVNNPKEMLQVDPQTKMYT 1380

1367 GVTGQVKSLLTSMYKEFLISSODGHQWTLFPONKVKYVFOGNDSTFYVNSLDPL 1426
 1367 GVTGQVKSLLTSMYKEFLISSODGHQWTLFPONKVKYVFOGNDSTFYVNSLDPL 1426

1381 GVTGQVKSLLTSMYKEFLISSODGHQWTLFPONKVKYVFOGNDSTFYVNSLDPL 1440
 1381 GVTGQVKSLLTSMYKEFLISSODGHQWTLFPONKVKYVFOGNDSTFYVNSLDPL 1440

1427 LFRYLRHPSQVHQAIRMEVLGCEAODLY 1457
 1427 LFRYLRHPSQVHQAIRMEVLGCEAODLY 1457

1441 LFRYLRHPSQVHQAIRMEVLGCEAODLY 1471
 1441 LFRYLRHPSQVHQAIRMEVLGCEAODLY 1471

RESULT 4
 ID R12971 standard; protein; 1440 AA.
 AC R12971.
 DE 02-OCT-1991 (first entry)
 DT Factor VIII:SQ.
 KW Factor VIII: B domain; haemophilia.
 PN MO9109122-A.
 PD 27-JUN-1991.
 PF 06-DEC-1990; SE0809.
 PR 15-DEC-1989; SE-004239.
 PA (KABI) KABI:VITROM AB.
 PI Almeted AB, Hellstrom EM, Larsson K, Lind P, Sandberg HI, Spira J, Sydow-Backman M;
 PI MPI: 91-208148/28.
 PT Recombinant human factor VIII deriv. deoxyribonucleic acid -
 PT encoding protein comprising two chains linked by segment of B domain.
 PS Disclosure; Fig 1; 35pp. English.
 CC The protein is a fusion between Phe 742 and Ser 1637 of the factor
 CC VIII protein (factor VIII:SQ). In order to produce a
 CC factor VIII deletion derivative that can be produced in vivo and/or
 CC in vitro, to a two chain protein consisting of polypeptide chains of
 CC 90 kD and 80 kD, the amino acid sequences surrounding Arg 740 and
 CC Arg 1648 have to be conserved in order to preserve the structural
 CC requirements for correct cleavage. In this example, amino acids 743
 CC to 1636 of the full-length factor VIII polypeptide are deleted. A
 CC new polypeptide chain is obtd. where there are 14 amino acids
 CC linking Arg 740 and Arg 1648. Of these 14 amino acids, the sequence
 CC of the five N-terminal ones directly corresponds to the five amino
 CC acids following Arg 740 in full-length factor VIII. Also, the sequence
 CC of the 12 C-terminal amino acids of the above 14 amino acids fragment
 CC directly corresponds to the 12 amino acids preceding Glu 1689 in full-
 CC length factor VIII, thus creating a 3 amino acid overlap between the
 CC N- and C-terminal regions of the B-domain.
 CC The factor VIII deriv. is useful for treating haemophilia or
 CC haemophilia A. It has the biological characteristics of plasma derived
 CC factor VIII.
 CC In order to index this example, the factor VIII:QD amino acid
 CC sequence was retrieved from W08800831 (P80265).
 CC The amino acid numbering in the above comments is reproduced from the
 CC fig. description in the specification. Note that Arg 740 is Arg 742

CC in the indexed sequence, etc. Also, Asn 745 (N-terminal link overlap)
CC is Asp 747 in P80265, but indexed as Asn to reproduce the fusion
CC fragment as shown in fig.1, and Asn 745 (C-terminal link overlap).
SQ Sequence 1440 AA:

Query Match 98.4%: Score 7668.5; DB 1; Length 1440;
Best Local Similarity 99.7%: Pred. No. 0;
Matches 1435; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 20 ATRRYLGAVELSDMYOSDGLGELPVDARPPRPVPSFPTSVYKKTLFEVETDHLFN 79
DB 2 ATRRYLGAVELSDMYOSDGLGELPVDARPPRPVPSFPTSVYKKTLFEVETDHLFN 61
QY 80 IAKRRPPMGLGPTIQAEVYDTVYITLKNNASHPVSLHAGVSTWAKSEGAEDDQTSQ 139
DB 62 IAKRRPPMGLGPTIQAEVYDTVYITLKNNASHPVSLHAGVSTWAKSEGAEDDQTSQ 121
QY 140 REKEDKVPFGSGSHYVQVLEKENGPMASDPLCLTYSLSHVDLYKDLNSGLIGALLVCR 199
DB 122 REKEDKVPFGSGSHYVQVLEKENGPMASDPLCLTYSLSHVDLYKDLNSGLIGALLVCR 181
QY 200 EGSIAKEKTLHKEFILFAVFEDEKSWHSETKNSLMODRDAASARAPKMHVNGYVNR 259
DB 182 EGSIAKEKTLHKEFILFAVFEDEKSWHSETKNSLMODRDAASARAPKMHVNGYVNR 241
QY 260 S-LGELICHRKSYWYVIGMGTPEVHSIFLEGHTFLVRNHRQASLEISPTFLTAOTL 318
DB 242 SLPLGLICHRKSYWYVIGMGTPEVHSIFLEGHTFLVRNHRQASLEISPTFLTAOTL 301
QY 319 LMDGOLFCHSHSHQDGEAYKVDSCPEEPOLRMKNNEADYDDDLDSMDVVR 378
DB 302 LMDGOLFCHSHSHQDGEAYKVDSCPEEPOLRMKNNEADYDDDLDSMDVVR 361
QY 379 FDDNPSPEFIOIRYAKKHPTWYHIAEEDMDYAPLVLPDDRYSKQTLNNGPQRI 438
DB 362 FDDNPSPEFIOIRYAKKHPTWYHIAEEDMDYAPLVLPDDRYSKQTLNNGPQRI 421
QY 439 GRKTKKRVFMAVYTDFTFTRALIOHESGILGPLYGEGVDLLIIFKNQASRPYNYPHG 498
DB 422 GRKTKKRVFMAVYTDFTFTRALIOHESGILGPLYGEGVDLLIIFKNQASRPYNYPHG 481
QY 499 ITDVRPLYSRRLPGVKVKKLKPFLIPGELIFKYKMTVTEDEPKSDPCLTRYSSFFNM 558
DB 482 ITDVRPLYSRRLPGVKVKKLKPFLIPGELIFKYKMTVTEDEPKSDPCLTRYSSFFNM 541
QY 559 ERDLASGLIGPLLCYKESVDQRGNQIMSDKRNVLFSVPDENRSMYLTENIQRLPMPA 618
DB 542 ERDLASGLIGPLLCYKESVDQRGNQIMSDKRNVLFSVPDENRSMYLTENIQRLPMPA 601
QY 619 GVQLEDPEFOASNMHSINGVYFDSLQSLVCLHEVAYWYILSIGAQTDPLSYFSGYTFK 678
DB 602 GVQLEDPEFOASNMHSINGVYFDSLQSLVCLHEVAYWYILSIGAQTDPLSYFSGYTFK 661
QY 679 HKMYVEDLTLFPSPSGEVFMSEMPGIMLIGCHNSDRNRKMTALLVSSCDKATGCTY 738
DB 662 HKMYVEDLTLFPSPSGEVFMSEMPGIMLIGCHNSDRNRKMTALLVSSCDKATGCTY 721
QY 739 EDJSEDISAVLLSKNNAIEPVSFSONPVLKRHOREITRTLOSQOEIDDDDTISVEMK 798
DB 722 EDJSEDISAVLLSKNNAIEPVSFSONPVLKRHOREITRTLOSQOEIDDDDTISVEMK 781
QY 799 KEDFDIYDEENOSRFSQKTRHYFAVERLMDYGMSSSPHYLRNRAQSGSVQPKFY 858
DB 782 KEDFDIYDEENOSRFSQKTRHYFAVERLMDYGMSSSPHYLRNRAQSGSVQPKFY 841
QY 859 VEOFTOSGFOPLYRGELNHLGLGYIRAEVDNITMTPRNQASRPYSYSSLIYE 918
DB 842 VEOFTOSGFOPLYRGELNHLGLGYIRAEVDNITMTPRNQASRPYSYSSLIYE 901
QY 919 EDORAGAPRRKRVFNPNTKTYFMVKVQHHMAPTKDEFDCKANAYSDVLEADVSGILIG 978
DB 902 EDORAGAPRRKRVFNPNTKTYFMVKVQHHMAPTKDEFDCKANAYSDVLEADVSGILIG 961

QY 979 PLVCHTNTLNPANGROYVQEPALFTIPEDETSMVFTENMERNCRAPCIMOEDPTFK 1038
DB 962 PLVCHTNTLNPANGROYVQEPALFTIPEDETSMVFTENMERNCRAPCIMOEDPTFK 1021
QY 1039 ENYRPHAINGYIMDPLPELVNAODORIRWYLLSMGNSNINSHSGVFTVRKKEEKM 1098
DB 1022 ENYRPHAINGYIMDPLPELVNAODORIRWYLLSMGNSNINSHSGVFTVRKKEEKM 1081
QY 1099 ALYNLYPGVEFTEMLPSKAGIMWRECLIGELHLAGMSTLFLVYSNKCOTPLGASGHIR 1158
DB 1082 ALYNLYPGVEFTEMLPSKAGIMWRECLIGELHLAGMSTLFLVYSNKCOTPLGASGHIR 1141
QY 1159 DFOITASGOYQOMAPKLARLHYSGSINAWSTKEPSPWIKVDLLAPMIIHGIKTGAROKF 1218
DB 1142 DFOITASGOYQOMAPKLARLHYSGSINAWSTKEPSPWIKVDLLAPMIIHGIKTGAROKF 1201
QY 1219 SLYTSOFITMSLDGKWQYRGRNSTGLTWFGNDSGKIHNIEMPIIARIIRLHP 1278
DB 1202 SLYTSOFITMSLDGKWQYRGRNSTGLTWFGNDSGKIHNIEMPIIARIIRLHP 1261
QY 1279 THYSIRSTLRMELMGCDLNSCSMPLGESKASISDAQITASSYFTMFAWSPSKARLHQ 1338
DB 1262 THYSIRSTLRMELMGCDLNSCSMPLGESKASISDAQITASSYFTMFAWSPSKARLHQ 1321
QY 1339 GRSNAMPQVNNPKEMLOVDYFQKTKVYGVTTQGVKSLTSMYKEFLISSQDGHQWL 1398
DB 1322 GRSNAMPQVNNPKEMLOVDYFQKTKVYGVTTQGVKSLTSMYKEFLISSQDGHQWL 1381
QY 1399 FFGNGKRVFQGNDSFTPVVNSLDPLRLTRYLRHPQSWHOLARMEVGECEADQLY 1457
DB 1382 FFGNGKRVFQGNDSFTPVVNSLDPLRLTRYLRHPQSWHOLARMEVGECEADQLY 1440

RESULT 5

P80265
ID P80265 standard; protein; 1516 AA.
AC P80265;
DT 10-OCT-1990 (first entry)
DE Modified factor VIII:C sequence with the Q744-D1563 deletion.
KW Modified factor VIII:C; maturation polypeptide; haemophilia;
KM blood coagulation; QD deletion.
OS Homo sapiens.
PN W08800831.A.
PD 11-FEB-1988.
PE 31-JUL-1987; U01814.
PR 01-AUG-1986; US-893375.
PA (BIOJ) Biogen NY (PASE/).
PI PASEK MP;
DR WPI: 88-049866/07.
DR N-PSDB: N80444.
PT New DNA sequences encoding modified factor VIII:C - with deletion of DNA
PT encoding maturation polypeptide, useful for high yield transformation.
PS Claim 3; Page 51-52-53-54; 97pp; English.
CC A major part of the sequence encoding the maturation polypeptide of
CC factor VIII:C is deleted, i.e. Gln 744 - Asp 1563. The QD deletion
CC retains approximately 90 amino acids of the maturation polypeptide
CC (four amino acids at the N-terminal end and 86 amino acids at
CC the C-terminal end). The full length factor VIII:C cDNA has two
CC changes with respect to the published sequence (EPO application 160457):
CC CTG to CTA at Leu 242 and TTC to CTC change at amino acid residue 1880
CC (Phe to Leu).
CC The product is produced in approx. 20 times higher
CC yields than previous recombinant produced factor VIII:C and are more
CC easily purified. The peptide is used for treating haemophilia A, both
CC acute and prolonged bleeding.
CC See also N80446 and N80447.
SQ Sequence 1516 AA:

Query Match 97.7%: Score 7620.5; DB 1; Length 1516;
Best Local Similarity 94.7%: Pred. No. 0;
Matches 1435; Conservative 1; Mismatches 2; Indels 77; Gaps 2;

```

QY 20 ATRRYLGAVELSMDIWDGELPVDARPPRPVPSFPFNISVYKKTLEFETDHLFN 79
DB 2 ATRRYLGAVELSMDIWDGELPVDARPPRPVPSFPFNISVYKKTLEFETDHLFN 61
QY 80 IAKPRPMWGLGPTIOAEVYDFVTITLKNMASHPSLAVGVSYKASGAEYDQTSQ 139
DB 62 IAKPRPMWGLGPTIOAEVYDFVTITLKNMASHPSLAVGVSYKASGAEYDQTSQ 121
QY 140 REKEDKVPFGSGHTYVWOLKENGPMASDPLCLITYSHYDVLVNDLSGLGALLVCR 199
DB 122 REKEDKVPFGSGHTYVWOLKENGPMASDPLCLITYSHYDVLVNDLSGLGALLVCR 181
QY 200 EGSIAEKQOTLHKFTLLFVPEDEGKSMHSETKNSLMODBASAPAMPKMTVNGYVR 259
DB 182 EGSIAEKQOTLHKFTLLFVPEDEGKSMHSETKNSLMODBASAPAMPKMTVNGYVR 241
QY 260 S-LPGLIGCHRSKVYVHVIOMGTTPREYHSIFLEGHTEFLVNHROASLEISPIITLAQTL 318
DB 242 SLIPGLIGCHRSKVYVHVIOMGTTPREYHSIFLEGHTEFLVNHROASLEISPIITLAQTL 301
QY 319 LMDLGOFLEFCHTISHOHODMEAYVAVDSCPEEPOLRMKNNERAEYDDDLTISEMDVVR 378
DB 302 LMDLGOFLEFCHTISHOHODMEAYVAVDSCPEEPOLRMKNNERAEYDDDLTISEMDVVR 361
QY 379 FDDDNPSFIOIRSVAKKHKPTWVHYTAAEEDMDYAPLVAPDDRSYKSOYLNGSPORI 438
DB 362 FDDDNPSFIOIRSVAKKHKPTWVHYTAAEEDMDYAPLVAPDDRSYKSOYLNGSPORI 421
QY 439 GRKYKRVFMAVYDEPKTRBAIOHESGLIGPLLYGEVGTLLIIFKNQASRPYNIYPHG 498
DB 422 GRKYKRVFMAVYDEPKTRBAIOHESGLIGPLLYGEVGTLLIIFKNQASRPYNIYPHG 481
QY 499 ITDVRPLYSRRLPGVXHLNDFPLPBEIFKYKWTVEEDGPTKSDRCITRYSSFFVM 558
DB 482 ITDVRPLYSRRLPGVXHLNDFPLPBEIFKYKWTVEEDGPTKSDRCITRYSSFFVM 541
QY 559 ERDIASGLIGPLLYCYKESVDQRCNOJMSDKRNVILFVSDEKNSWYLTENIOFLEPNA 618
DB 542 ERDIASGLIGPLLYCYKESVDQRCNOJMSDKRNVILFVSDEKNSWYLTENIOFLEPNA 601
QY 619 GVOLEDEPEFOASNTMHSINGVPEFSLDLYCLHVAAYWILSICAGODFLSVFSSGTFK 678
DB 602 GVOLEDEPEFOASNTMHSINGVPEFSLDLYCLHVAAYWILSICAGODFLSVFSSGTFK 661
QY 679 HKAVYEDTLTFPPSGETVFMENPGMLILGCHNSDFRNKMTALLKVSSCDKNTGDY 738
DB 662 HKAVYEDTLTFPPSGETVFMENPGMLILGCHNSDFRNKMTALLKVSSCDKNTGDY 721
QY 739 EDSDYEDISAVLLSKNNALIEPRSF----- 762
DB 722 EDSDYEDISAVLLSKNNALIEPRSF----- 781
QY 762 -----SONPPLKXHOEIRRTLOS 782
DB 782 ILISINACESNHAIAINEGOKPELEVYMAKOGTERELCSQNPVYLKXHOEIRRTLOS 841
QY 783 DOEIEDYDITISVEKKEDFDIYDENQSPRSFOKTRHYFAIAVERLMDYGNSSSPHV 842
DB 842 DOEIEDYDITISVEKKEDFDIYDENQSPRSFOKTRHYFAIAVERLMDYGNSSSPHV 901
QY 843 LNRNAGSGVPOFKKVVFOEFTDGSFTOPLYRGELEHGLGPIYRAVEDNIMVTFRN 902
DB 902 LNRNAGSGVPOFKKVVFOEFTDGSFTOPLYRGELEHGLGPIYRAVEDNIMVTFRN 961
QY 903 QASRPSYSSLSIYEEOROGAEPKRNFKVKNETKYFMKVOHMAETKDEFCAMAY 962
DB 962 QASRPSYSSLSIYEEOROGAEPKRNFKVKNETKYFMKVOHMAETKDEFCAMAY 1021
QY 963 FSDVDLEKDVHSGLIGPLLYCHTNTLNPAHGRQVTVQAEFAFTTIFDETCKSWYFENNER 1022
DB 1022 FSDVDLEKDVHSGLIGPLLYCHTNTLNPAHGRQVTVQAEFAFTTIFDETCKSWYFENNER 1081

```

```

QY 1023 NCRAPCNIOMEDPYFKENYRERHAIINGYIMDTLPGLVVAODORIRWYLLSGNSNENHSIH 1082
DB 1082 NCRAPCNIOMEDPYFKENYRERHAIINGYIMDTLPGLVVAODORIRWYLLSGNSNENHSIH 1141
QY 1083 FSGHYFYRKKEEKEMALXNYIPGVFTEVLELPSKAGIMVECLIGHLHAGNSTLEFLY 1142
DB 1142 FSGHYFYRKKEEKEMALXNYIPGVFTEVLELPSKAGIMVECLIGHLHAGNSTLEFLY 1201
QY 1143 SNKCOTPLGMSAGHIRDFOITASGOYOMAPKLARLHYSGSINAMSTKEPFSWIKVDLLA 1202
DB 1202 SNKCOTPLGMSAGHIRDFOITASGOYOMAPKLARLHYSGSINAMSTKEPFSWIKVDLLA 1261
QY 1203 PMIHGITQAGARQKFSGLYISQFIMTSLDGKKWQYIRGSTSTLWVFRNVDSSGIXH 1262
DB 1262 PMIHGITQAGARQKFSGLYISQFIMTSLDGKKWQYIRGSTSTLWVFRNVDSSGIXH 1321
QY 1263 NIFNPILARIYRLHPHTYSIRSTLRMELMCDLNSCMLPGMESKASIDAOITASSYFT 1322
DB 1322 NIFNPILARIYRLHPHTYSIRSTLRMELMCDLNSCMLPGMESKASIDAOITASSYFT 1381
QY 1323 NMFATWSPSKARLHLQGRSNAMRPVNNPKEMLYVDFQKTKRVTVGTTQGVKSILTSMTY 1382
DB 1382 NMFATWSPSKARLHLQGRSNAMRPVNNPKEMLYVDFQKTKRVTVGTTQGVKSILTSMTY 1441
QY 1383 KEFLISSODGHQWTLFQNGKVKYFQGNOSFTPVVNSLDPLLTRYLRIHPQSWHIOI 1442
DB 1442 KEFLISSODGHQWTLFQNGKVKYFQGNOSFTPVVNSLDPLLTRYLRIHPQSWHIOI 1501
QY 1443 ALRMEVYLCEADOLY 1457
DB 1502 ALRMEVYLCEADOLY 1516

```

RESULT 6
 ID P80267
 AC P80267;
 DT 10-OCT-1990 (first entry)
 KW Modified factor VIII:C sequence with the R740-E1649 deletion.
 DE Modified factor VIII:C; maturation polypeptide; haemophilia;
 KM blood coagulation; RE deletion.
 OS Homo sapiens.
 PN MO8800831-A.
 PD 11-FEB-1988.
 PF 31-JUL-1987; U01814.
 PR 01-AUG-1986; US-893375.
 PA (Bio) Biogen NV (PASE/).
 PI PASEK MP;
 DR WPI; 88-049866/07.
 DR N-PSDB; N80446.
 PT New DNA sequences encoding modified factor VIII:C - with deletion of DNA
 encoding maturation polypeptide, useful for high yield transformation.
 PS Claim 3; Page 57-58-59-60; 97bp; English.
 CC The entire sequence encoding the maturation polypeptide of
 CC factor VIII:C is deleted, i.e. Arg 740-Glu 1649.
 CC The full length factor VIII:C cDNA has two changes with respect to the
 CC published sequence (BPO application 160457):
 CC CAG to CTA at Leu 242 and TMC to CTC change at amino acid residue 1880
 CC (Phe to Leu). The product is produced in approx. 20 times higher
 CC yields than previous recombinant produced factor VIII:C and are more
 CC easily purified. The peptide is used for treating haemophilia A, both
 CC acute and prolonged bleeding.
 CC See also N80444 and N80447.
 SQ Sequence 1425 AA;

Query Match 97.3%; Score 7590; DB 1; Length 1425;
 Best Local Similarity 98.9%; Pred. No. 0;
 Matches 1422; Conservative 0; Mismatches 2; Indels 14; Gaps 1;

```

QY 20 ATRRYLGAVELSMDIWDGELPVDARPPRPVPSFPFNISVYKKTLEFETDHLFN 79
DB 2 ATRRYLGAVELSMDIWDGELPVDARPPRPVPSFPFNISVYKKTLEFETDHLFN 61

```

```

QY 80 IAKRPPMGLGPTIOAEVYDVTYITLKNNASHPVSLHAGVSYWKASBGAEDDQTSQ 139
DB 62 IAKRPPMGLGPTIOAEVYDVTYITLKNNASHPVSLHAGVSYWKASBGAEDDQTSQ 121
QY 140 REKDDKVFPGSGHTYVQVLEKNGPMASDPLCLTYSLSHVDLVKDNLSGLIGALLVCR 199
DB 122 REKDDKVFPGSGHTYVQVLEKNGPMASDPLCLTYSLSHVDLVKDNLSGLIGALLVCR 181
QY 200 EGSIAKETOTLHKFILLFAVFDGKSMHSETKNSLMODRDAASARAPKHHYNGVYNR 259
DB 182 EGSIAKETHTLHKFILLFAVFDGKSMHSETKNSLMODRDAASARAPKHHYNGVYNR 241
QY 260 SLPLICGHRKSVYWHYIGMGTPEVHSIFLEGHTFLVNRHQAQSLSEIPTFLTAOTLL 319
DB 242 SLPLICHRKSVYWHYIGMGTPEVHSIFLEGHTFLVNRHQAQSLSEIPTFLTAOTLL 301
QY 320 MDLQGFLLFCHSISSHODGMEAYVVDSCPEEPQLRMKNNEADYDDDLTDSMDVYRF 379
DB 302 MDLQGFLLFCHSISSHODGMEAYVVDSCPEEPQLRMKNNEADYDDDLTDSMDVYRF 361
QY 380 DDDNSPFIQIRSVAKKHPTWVHYIAAEEEDMDYAPLYLAPDDRYSKSOYLNNQPQRIQ 439
DB 362 DDDNSPFIQIRSVAKKHPTWVHYIAAEEEDMDYAPLYLAPDDRYSKSOYLNNQPQRIQ 421
QY 440 KKYKVRPMAYTDETFKTRALIQHESGILGPLYGEGDILLIFKNQASRPYIYPHGI 499
DB 422 KKYKVRPMAYTDETFKTRALIQHESGILGPLYGEGDILLIFKNQASRPYIYPHGI 481
QY 500 TDVPLVSRRLPKGVKHLKDFPLPGFEFKKMYVVEDGTSKDPRLTLTYSSFFVME 559
DB 482 TDVPLVSRRLPKGVKHLKDFPLPGFEFKKMYVVEDGTSKDPRLTLTYSSFFVME 541
QY 560 RDLASGLIGPLLYCKESVDQGNQIMSDKRNVLISVFEDNRSWYLTENIQRELPNAG 619
DB 542 RDLASGLIGPLLYCKESVDQGNQIMSDKRNVLISVFEDNRSWYLTENIQRELPNAG 601
QY 620 VQLEDPQOASINMHSINGYVFDLSQISVCLHEVAWYIILSIGQDTLISVFESGYTFKH 679
DB 602 VQLEDPQOASINMHSINGYVFDLSQISVCLHEVAWYIILSIGQDTLISVFESGYTFKH 661
QY 680 KMYVEDTLTLPFSGEYFVEMENMGILGCHNSDFNRMTALLKSSCDKNTGDYE 739
DB 662 KMYVEDTLTLPFSGEYFVEMENMGILGCHNSDFNRMTALLKSSCDKNTGDYE 721
QY 740 DSYEDISAYLLSKNNALERPFSFSONPVLKROHEITRTTLOSQOEIEDYDITISVEKK 799
DB 722 DSYEDISAYLLSKNNALERPFSFSONPVLKROHEITRTTLOSQOEIEDYDITISVEKK 767
QY 800 EDFDIYDEDENQSPRSFOKTRRHFIYAVERLMDYGMSSSPHVLNRNQAQSGSVDFOKKV 859
DB 768 EDFDIYDEDENQSPRSFOKTRRHFIYAVERLMDYGMSSSPHVLNRNQAQSGSVDFOKKV 827
QY 860 FOEFTDGSFTQPLVRGELNEHLGLGPYIRAEVEDNIMVTRNQAQSRYSYSSISTEE 919
DB 828 FOEFTDGSFTQPLVRGELNEHLGLGPYIRAEVEDNIMVTRNQAQSRYSYSSISTEE 887
QY 920 DOROGAERKMFVAPNETKITFMKVOHMAPTKDEFDCKANAFFSDVLEKDVHSGILGP 979
DB 888 DOROGAERKMFVAPNETKITFMKVOHMAPTKDEFDCKANAFFSDVLEKDVHSGILGP 947
QY 980 LLVCHTNTLNAHGRQVTVQEPALFTTIFDETksWYFTENNERNCRACNIOEDPTFE 1039
DB 948 LLVCHTNTLNAHGRQVTVQEPALFTTIFDETksWYFTENNERNCRACNIOEDPTFE 1007
QY 1040 NYRFAINGYIMDTLPGVYMAODORIRRYLLSMGSNENIHSHGSHFYTRKKEEYMA 1099
DB 1008 NYRFAINGYIMDTLPGVYMAODORIRRYLLSMGSNENIHSHGSHFYTRKKEEYMA 1067
QY 1100 LYNLYPGVFETVEMLPKAGIWRVCLLIGELHAGMSTLFLVYSNRKOTPLGMAAGHTRD 1159
DB 1068 LYNLYPGVFETVEMLPKAGIWRVCLLIGELHAGMSTLFLVYSNRKOTPLGMAAGHTRD 1127

```

```

QY 1160 FOITASGOYQOMAPKLARLRYSGSINAMSTKEPSPKIVULLAPMITHGIKTOGAROKES 1219
DB 1128 FOITASGOYQOMAPKLARLRYSGSINAMSTKEPSPKIVULLAPMITHGIKTOGAROKES 1187
QY 1220 SLVYSOFIIMSLDGKWKQOTYRGNSTGTLMVFGNVSSGIGKHIFNPPIIARIIRLHPT 1279
DB 1188 SLVYSOFIIMSLDGKWKQOTYRGNSTGTLMVFGNVSSGIGKHIFNPPIIARIIRLHPT 1247
QY 1280 HYSIRSLRMLMGCDLNSCSMPLGMEKASISDAQTASSYFTNMFTWSPSKARLHOG 1339
DB 1248 HYSIRSLRMLMGCDLNSCSMPLGMEKASISDAQTASSYFTNMFTWSPSKARLHOG 1307
QY 1340 RSNAMRPQVNNPKFEMLOVDQKTKMYTGVTTQGYKSLITSMYKKEFLISSODGHQWTLF 1399
DB 1308 RSNAMRPQVNNPKFEMLOVDQKTKMYTGVTTQGYKSLITSMYKKEFLISSODGHQWTLF 1367
QY 1400 FONGKVFVFOGNODSFYPVNSLDPPLTFLYRLRHPOSWHOLRREVJGCEADODLY 1457
DB 1368 FONGKVFVFOGNODSFYPVNSLDPPLTFLYRLRHPOSWHOLRREVJGCEADODLY 1425

```

RESULT 7

```

ID P80268 standard: protein; 1424 AA.
AC P80268;
DT 10-OCT-1990 (first entry)
DE Modified factor VIII:C sequence with the R740-D1658 deletion.
KM Modified factor VIII:C; maturation polypeptide; haemophilia;
KW blood coagulation; RD deletion; procoagulant.
OS Homo sapiens.
PN W08800831-A.
PD 11-FEB-1988.
PF 31-JUL-1987; 001814.
PR 01-AUG-1986; US-893375.
PA (BIOJ) Biogen NY (PASE/).
PI PASEK MP.
DR WPI; 88-049866/07.
DR N-PSDE; N80447.
PT New DNA sequences encoding modified factor VIII:C - with deletion of DNA
PT encoding maturation polypeptide, useful for high yield transformation.
PS Claim 3: Page 60-61-62-63; 97pp; English.
CC The RD deletion removes the DNA from Ser 741 to Ser 1657.
CC A major part of the sequence encoding the maturation polypeptide of
CC factor VIII:C is deleted, i.e. Gln 744 - Asp 1657.
CC The full length Factor VIII:C cDNA has two changes with respect to the
CC published sequence (EPO application 160457):
CC CNG to CTA at Leu 242 and Trp to CTrp change at amino acid residue 1880
CC (The to Leu). The product is produced in approx. 20 times higher
CC yields than previous recombinant produced factor VIII:C and are more
CC easily purified. The peptide is used for treating haemophilia A, both
CC acute and prolonged bleeding.
CC See also N80444 and N80446.
SQ Sequence 1424 AA.

```

Query Match

Best Local Similarity 98.8%; Score 7587; DB 1; Length 1424;

Matches 142; Conservative 1; Mismatches 2; Indels 14; Gaps 1;

```

QY 20 ATRRYIGAVELSDMYQSDLGELPVDAKPPRPVKSPFRTSVYKTLFVEETDHLFN 79
DB 1 ATRRYIGAVELSDMYQSDLGELPVDAKPPRPVKSPFRTSVYKTLFVEETDHLFN 60
QY 80 IAKRPPMGLGPTIOAEVYDVTYITLKNNASHPVSLHAGVSYWKASBGAEDDQTSQ 139
DB 61 IAKRPPMGLGPTIOAEVYDVTYITLKNNASHPVSLHAGVSYWKASBGAEDDQTSQ 120
QY 140 REKDDKVFPGSGHTYVQVLEKNGPMASDPLCLTYSLSHVDLVKDNLSGLIGALLVCR 199
DB 121 REKDDKVFPGSGHTYVQVLEKNGPMASDPLCLTYSLSHVDLVKDNLSGLIGALLVCR 180
QY 200 EGSIAKETOTLHKFILLFAVFDGKSMHSETKNSLMODRDAASARAPKHHYNGVYNR 259
DB 1068 EGSIAKETOTLHKFILLFAVFDGKSMHSETKNSLMODRDAASARAPKHHYNGVYNR 1127

```

```

Db 181 EGSIAKETHTLHKFTLLFAVFEDEKSMHSETKNSLMODROAASARAMPKMTHTVNGYVR 240
QY 260 SLPLGLGCHRRKSVYWHVIGMGTTPPEVHSIFLEGHTFLVNRHQAASLEISPTIFLTAQTL 319
Db 241 SLPLGLGCHRRKSVYWHVIGMGTTPPEVHSIFLEGHTFLVNRHQAASLEISPTIFLTAQTL 300
QY 320 MDLGFLLFCHSSHQHDGMEAYVAVDSCPEEPOLRMKNNEAEYDDDLTJSEMDVYRF 379
Db 301 MDLGFLLFCHSSHQHDGMEAYVAVDSCPEEPOLRMKNNEAEYDDDLTJSEMDVYRF 360
QY 380 DDNSPSFQIISVAKKHKHTVWHYIAAEEDMDYAPVLAPDDRYSQYLNNNGPORG 439
Db 361 DDNSPSFQIISVAKKHKHTVWHYIAAEEDMDYAPVLAPDDRYSQYLNNNGPORG 420
QY 440 RYKXKVRMAAYDETPKTRERAIQHSGLIGPLLYGEVDTLLIFKKNQASRYNTIYPHGI 499
Db 421 RYKXKVRMAAYDETPKTRERAIQHSGLIGPLLYGEVDTLLIFKKNQASRYNTIYPHGI 480
500 TDVRLYSRLKGYKHLKDFLLGEIPEKRYKMTVVEDGPTKSDPCLTRYSSFFVNM 559
481 TDVRLYSRLKGYKHLKDFLLGEIPEKRYKMTVVEDGPTKSDPCLTRYSSFFVNM 540
QY 560 RDLASGLIGPLLYCKESVDQGNQIMSDKRVNLFSEVDENRSMWLTENIQRFLLPNPAG 619
Db 541 RDLASGLIGPLLYCKESVDQGNQIMSDKRVNLFSEVDENRSMWLTENIQRFLLPNPAG 600
QY 620 VOLEDEPFQASNMHSINQVVDLSLQSLCHLEVAWYLLSTGAQDFELSVFSGTETH 679
Db 601 VOLEDEPFQASNMHSINQVVDLSLQSLCHLEVAWYLLSTGAQDFELSVFSGTETH 660
QY 680 KMYEEDTLLEFPSEGTVMSEMPGLMILGCHNSDFRNRGMTALIKVSSCCNTGDYVE 739
Db 661 KMYEEDTLLEFPSEGTVMSEMPGLMILGCHNSDFRNRGMTALIKVSSCCNTGDYVE 720
QY 740 DSYEDIISAVLLSKNNAIEPRSEQNPVYLRHOREITRTTLOSDEIDYDTISYEMKK 799
Db 721 DSYEDIISAVLLSKNNAIEPRSEQNPVYLRHOREITRTTLOSDEIDYDTISYEMKK 766
QY 800 EDDDIYDEENOSPRSFQKTRHYFAAVERLDYGMSSPHYLKRAAGSGVPOKRYV 859
Db 767 EDDDIYDEENOSPRSFQKTRHYFAAVERLDYGMSSPHYLKRAAGSGVPOKRYV 826
QY 880 FQETDGSFTQPLRYRGLNHLGLPYIRAEVEDNIMVTFRNQASRPYSFSSLSIYZEE 919
Db 827 FQETDGSFTQPLRYRGLNHLGLPYIRAEVEDNIMVTFRNQASRPYSFSSLSIYZEE 886
920 DQNGAEPKRNFPKPNETKYEFKQVQHMAPTDEDFCKAMAFSDVDLEKDVHSGILGP 979
887 DQNGAEPKRNFPKPNETKYEFKQVQHMAPTDEDFCKAMAFSDVDLEKDVHSGILGP 946
QY 980 LLYCHNTLTPAAGROVTOEFALFTIEDETKSWYFTENNERNCRAFCNIQMEDPTFKE 1039
Db 947 LLYCHNTLTPAAGROVTOEFALFTIEDETKSWYFTENNERNCRAFCNIQMEDPTFKE 1006
QY 1040 NYRPHAINQYIMDTLPGLVNAQDQIRIWMYLLSMGSENHSHIFSGHVFYRKKKEEKMA 1099
Db 1007 NYRPHAINQYIMDTLPGLVNAQDQIRIWMYLLSMGSENHSHIFSGHVFYRKKKEEKMA 1066
QY 1100 LYNLYPGEVETVEMLPKAGIWRRECLIGELHAGMSTFLVYSNKCOTPLGASGHIR 1159
Db 1067 LYNLYPGEVETVEMLPKAGIWRRECLIGELHAGMSTFLVYSNKCOTPLGASGHIR 1126
QY 1160 FQITASGOYQOMAPKLARLHYSGSINAMSTKEPFSWIKVDLLAPMIHGIKTQGAROKFS 1219
Db 1127 FQITASGOYQOMAPKLARLHYSGSINAMSTKEPFSWIKVDLLAPMIHGIKTQGAROKFS 1186
QY 1220 SLVTSQFTIMYSLDGKQWQYIRGNSTGLTAWFGVNDSSGCKHNIENPILARIIRHPT 1279
Db 1187 SLVTSQFTIMYSLDGKQWQYIRGNSTGLTAWFGVNDSSGCKHNIENPILARIIRHPT 1246
QY 1280 HYSIRSLRLMELMGCDLNSCSMPLGMSKASIDAOITASSFTMFWTSPSKARLHOG 1339
Db 1247 HYSIRSLRLMELMGCDLNSCSMPLGMSKASIDAOITASSFTMFWTSPSKARLHOG 1306

```

```

QY 1340 RSNARPOVNNPKEMLOVDFOKTMKVTGVTTOGVKSLTSMYKEFLISSQDGHQWTLF 1399
Db 1307 RSNARPOVNNPKEMLOVDFOKTMKVTGVTTOGVKSLTSMYKEFLISSQDGHQWTLF 1366
QY 1400 FONGKRVFQGNODSFTPVVNSLDPILTRYLRHPOSWWHQIALRMEVIGCAADLY 1457
Db 1367 FONGKRVFQGNODSFTPVVNSLDPILTRYLRHPOSWWHQIALRMEVIGCAADLY 1424

RESULT 8
P91169
ID P91169 standard; protein; 1424 AA.
AC P91169;
DT 26-JUN-1990 (first entry)
DE Sequence of 740 Arg-1649 Glu human Factor VIII:C
KW Human Factor VIII:C; Ad.RE.neo; 740 Arg-1649 Glu Factor VIII:C;
KW haemophilia A.
OS Homo sapiens.
PN EP-306968-A.
PD 15-MAR-1989.
PE 09-SEP-1988; 114769.
PR 08-APR-1988; JP-085454.
RA (KAGA) Chemo-Sero-Therap (Teij).
RA Sugiyama T, Masuda K, Tajima Y, Yonemura H;
PI WPI; 89-078467/11.
DR N-PSDB; N90654.
PT Prodn. of recombinant human Factor-VIII-C -
PT using animal cells transformed with a vector contg. the gene for
PT Factor VIII:C and a promoter
PS Figure 1(1) - 1(13); 32pp; English.
CC Arg-740 of the carboxyl terminus of the H chain is
CC directly bonded by a peptide bond to Glu-1649 of the amino terminus of
CC l chain. A pref. expression vector used to transform animal cell so
CC that they produce human Factor VIII:Cis plasmid Ad.RE.neo.
CC The expression vector has at least one promoter upstream of n90654.
CC The transformants can constantly and continuously produce human Factor
CC VIII:C in high yield on a commercial scale. The human Factor VIII:C so
CC produced is considered to corresp. to the smallest species of active and
CC intact Factor VIII:C molecules in the human blood plasma. It is useful
CC for treating haemophilia A patients.
SQ Sequence 1424 AA;

Query Match 97.3%; Score 7583; DB 1; Length 1424;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1421; Conservative 0; Mismatches 3; Indels 14; Gaps 1;

QY 20 ATRRYLGAVELSMQYMSDGLPVDARPPRPVKSFPFNTSVYKKTLEVEFTDHLFN 79
Db 1 ATRRYLGAVELSMQYMSDGLPVDARPPRPVKSFPFNTSVYKKTLEVEFTDHLFN 60
QY 80 IAKPRPMMGLGPTIOAEVYDTVVITLKMAASHPVSLHAGVSYMKASEGAEXDDQTSQ 139
Db 61 IAKPRPMMGLGPTIOAEVYDTVVITLKMAASHPVSLHAGVSYMKASEGAEXDDQTSQ 120
QY 140 REKEDDKVFGSGSHYVWQVYLKENGPMASDPLCTLYSLSHVDLYKVDLNSGLIGALLYCR 199
Db 121 REKEDDKVFGSGSHYVWQVYLKENGPMASDPLCTLYSLSHVDLYKVDLNSGLIGALLYCR 180
QY 200 EGSIAKETHTLHKFTLLFAVFEDEKSMHSETKNSLMODROAASARAMPKMTHTVNGYVR 259
Db 181 EGSIAKETHTLHKFTLLFAVFEDEKSMHSETKNSLMODROAASARAMPKMTHTVNGYVR 240
QY 260 SLPLGLGCHRRKSVYWHVIGMGTTPPEVHSIFLEGHTFLVNRHQAASLEISPTIFLTAQTL 319
Db 241 SLPLGLGCHRRKSVYWHVIGMGTTPPEVHSIFLEGHTFLVNRHQAASLEISPTIFLTAQTL 300
QY 320 MDLGFLLFCHSSHQHDGMEAYVAVDSCPEEPOLRMKNNEAEYDDDLTJSEMDVYRF 379
Db 301 MDLGFLLFCHSSHQHDGMEAYVAVDSCPEEPOLRMKNNEAEYDDDLTJSEMDVYRF 360
QY 380 DDNSPSFQIISVAKKHKHTVWHYIAAEEDMDYAPVLAPDDRYSQYLNNNGPORG 439

```

Db 361 DDNDSFQIOTSVAKKHKTWVHTAAEEDMDYAPLYLAHDNSYKQYLNNGFORIG 420
Qy 440 RRYKAVRNAYTDEFTKTRREALIOHESGILGLLYGEVGDTLIIIFKNASRPNIYPHGI 499
Db 421 RRYKAVRNAYTDEFTKTRREALIOHESGILGLLYGEVGDTLIIIFKNASRPNIYPHGI 480
Qy 500 TVVRPLYSRLKGVKHLKDFPLILPEIKRYKWTYVDEGPPKSPDRCLTRYSSVYVNE 559
Db 481 TVVRPLYSRLKGVKHLKDFPLILPEIKRYKWTYVDEGPPKSPDRCLTRYSSVYVNE 540
Qy 560 RDLASGLIPLILICYKESVDQGNQIMSDKRNVIIFSVYDENRSMYLFENIORFLPNPAG 619
Db 541 RDLASGLIPLILICYKESVDQGNQIMSDKRNVIIFSVYDENRSMYLFENIORFLPNPAG 600
Qy 620 VOLEDEPFOASIMHSINGVYEDSLQSVCLHEVAYWYLLISGAQTDLFVSEFSGYTFKH 679
Db 601 VOLEDEPFOASIMHSINGVYEDSLQSVCLHEVAYWYLLISGAQTDLFVSEFSGYTFKH 660
Qy 680 KAVYEDTLTLFPFSGEYVMSHNPGLMILGCHNSDFNRKGMATILIKVSSCDKNTGDYVE 739
Db 661 KAVYEDTLTLFPFSGEYVMSHNPGLMILGCHNSDFNRKGMATILIKVSSCDKNTGDYVE 720
Qy 740 DSYEDISAYLLSKNNAIEPRFSQNPVYAKRHQRETTTLTSDOEIDYDITISYEMK 799
Db 721 DSYEDISAYLLSKNNAIEPRFSQNPVYAKRHQRETTTLTSDOEIDYDITISYEMK 766
Qy 800 EDFDIYDEENOSPRSFQKTRHYFAAVERLMDYGMSSPHYLNRRAQSGSVPOKRYV 859
Db 767 EDFDIYDEENOSPRSFQKTRHYFAAVERLMDYGMSSPHYLNRRAQSGSVPOKRYV 826
Qy 860 FOEFIDGSTOPLYKGLNEHGLGPTIRAEVDNIMYTFNRQASRPYSSYSLISYEB 919
Db 827 FOEFIDGSTOPLYKGLNEHGLGPTIRAEVDNIMYTFNRQASRPYSSYSLISYEB 886
Qy 920 DQROGAEPKRNKVPKMETKTYFMKVQHMAPTKDEDFCAMAIFYSDVLEKDYHSLIIP 979
Db 887 DQROGAEPKRNKVPKMETKTYFMKVQHMAPTKDEDFCAMAIFYSDVLEKDYHSLIIP 946
Qy 980 LLYCHTNTLNPAGROVYVOEFAFLPTIFDETCKSWYFTEENMERNCRAPNIOEMDETFKE 1039
Db 947 LLYCHTNTLNPAGROVYVOEFAFLPTIFDETCKSWYFTEENMERNCRAPNIOEMDETFKE 1006
Qy 1040 NRYFAINCYIMDTLPGLVMAODQIRMTYLLSNGSNENHSHFSGHVFTVRKKEEYKMA 1099
Db 1007 NRYFAINCYIMDTLPGLVMAODQIRMTYLLSNGSNENHSHFSGHVFTVRKKEEYKMA 1066
Qy 1100 LYNLPGVEYETVEMLPKSKAGIRVECLIGEHLAGMSTLFVYYSNKCQPTPLGMAASHID 1159
Db 1067 LYNLPGVEYETVEMLPKSKAGIRVECLIGEHLAGMSTLFVYYSNKCQPTPLGMAASHID 1126
Qy 1160 FOITASGOYGOMAPKLARLHAGSINAMSTKEPFWIKYDLAPMIIHGIKTGAGAROKFS 1219
Db 1127 FOITASGOYGOMAPKLARLHAGSINAMSTKEPFWIKYDLAPMIIHGIKTGAGAROKFS 1186
Qy 1220 SLYISOFIIMYSLDGKMQTYRGNSGTGLMVEFGVNDSSGIKHNINPPIIARIYIRLHP 1279
Db 1187 SLYISOFIIMYSLDGKMQTYRGNSGTGLMVEFGVNDSSGIKHNINPPIIARIYIRLHP 1246
Qy 1280 HYSIRSLTLMELMGDLNCSMPLGMEKRAISDAQITASSYTFNMATSPSKARLHLQ 1339
Db 1247 HYSIRSLTLMELMGDLNCSMPLGMEKRAISDAQITASSYTFNMATSPSKARLHLQ 1306
Qy 1340 RSNAMPQVNNPREMLQVFOKTMKTGYTQGVKSLSMYKEPFLISSODGHOHWTF 1399
Db 1307 RSNAMPQVNNPREMLQVFOKTMKTGYTQGVKSLSMYKEPFLISSODGHOHWTF 1366
Qy 1400 FONGKAVYFQGNDSFTPVVNSLDPLLTTRYLRHPOQSVHQAIALMEVYLGCEAODLY 1457
Db 1367 FONGKAVYFQGNDSFTPVVNSLDPLLTTRYLRHPOQSVHQAIALMEVYLGCEAODLY 1424

W18670
ID W18670 standard; Protein; 1661 AA.
AC W18670;
DT 10-AUG-1997 (first entry)
DE Factor VIII-dB695-HCII.
KW Factor VIII-dB695-HCII; heparin cofactor II; blood coagulation;
KM blood clotting; procoagulant; anticoagulant; antithrombotic;
KN haemophilia; gene therapy.
OS Homo sapiens.
FH Key Location/Qualifiers
FT domain 731..760
FT /label= HCII
FT /note= "heparin cofactor II acidic domain"
FN W0918315-A1.
PD 22-MAY-1997.
PE 13-NOV-1996; E04977.
PR 13-NOV-1995; US-558107.
PA (IMMO) IMMUNO AG.
PI Vooberg J;
DR WPI; 97-289291/26.
DR N-PSDB; T69811.
PT Hybrid Factor VIII with modified activity, comprises region from
PT donor anticoagulant or antithrombotic protein - useful for treatment
PT of coagulation disorders
PS Claim 11; Page 52-60; 96pp; English.
CC Factor VIII-dB695-HCII (W18670) is a hybrid protein in which amino
CC acids 712-736 of Factor-dB695 (Factor VIII del868-1562) B-domain
CC are replaced by amino acids 51-80 from the acidic region (and
CC potential thrombin-binding site) of human heparin cofactor II
CC (HCII). It is the expression product of Factor VIII-dB695-HCII
CC DNA (T69811) in plasmid pCUB-dB695-HCII. The hybrid protein, which
CC can be expressed using gene therapy techniques, has increased
CC procoagulant activity owing to the HCII acidic region, and can be
CC used to treat blood coagulation disorders such as haemophilia A.
SQ Sequence 1661 AA;

Query Match 96.8%; Score 7544; DB 1; Length 1661;
Best Local Similarity 86.2%; Pred. No. 0;
Matches 1437; Conservative 5; Mismatches 9; Indels 216; Gaps 4;

Qy 1 MHEIISTCFCLLAFPCSATIRRYLLGAVELSWDMQSDGLGLPVDARPPRPVPSFPN 60
Db 1 MHEIISTCFCLLAFPCSATIRRYLLGAVELSWDMQSDGLGLPVDARPPRPVPSFPN 60
Qy 61 TSVYVKKTLFVFTLHFNIAKPRPMGLGPTIOAEYDTVVITLKNMASHPSLAHV 120
Db 61 TSVYVKKTLFVFTLHFNIAKPRPMGLGPTIOAEYDTVVITLKNMASHPSLAHV 120
Qy 121 GVSYKASBGAEYDQTSQREKEDKVPFGSGHTYVWYVLKENGPMASDPLCLTYSYLH 180
Db 121 GVSYKASBGAEYDQTSQREKEDKVPFGSGHTYVWYVLKENGPMASDPLCLTYSYLH 180
Qy 181 VDLVYDNLNGLIGALLVCREGSLAKEKTOYLLKFTLLFVAVDEGKSMHSETNLSMODRD 240
Db 181 VDLVYDNLNGLIGALLVCREGSLAKEKTOYLLKFTLLFVAVDEGKSMHSETNLSMODRD 240
Qy 241 AASARAMPMTHTVNVNRSPLGLGCHRSKYWYVIGMTTPEVHSIFLEBHTFLVRH 300
Db 241 AASARAMPMTHTVNVNRSPLGLGCHRSKYWYVIGMTTPEVHSIFLEBHTFLVRH 300
Qy 301 ROASLEISPTFLTQOTLLMDGQFLFCHISSHODGMEAVYKVDSCPEEPQLMKNE 360
Db 301 ROASLEISPTFLTQOTLLMDGQFLFCHISSHODGMEAVYKVDSCPEEPQLMKNE 360
Qy 361 EAEYDDDDLTDEEMVYVRFDDNDSPTIOIRSVAKKHPTWVHTAAEEDMDYAPLYLA 420
Db 361 EAEYDDDDLTDEEMVYVRFDDNDSPTIOIRSVAKKHPTWVHTAAEEDMDYAPLYLA 420
Qy 421 PDDRYSKQYLNNQFORIGRKYKVFMAAYTDETKTRTREAIOHESGILGLLYGEVGD 480
Db 421 PDDRYSKQYLNNQFORIGRKYKVFMAAYTDETKTRTREAIOHESGILGLLYGEVGD 480

VIII, along with calcium and phospholipid, acts as a cofactor for factor IXa, when it converts factor X to the activated form (factor Xa). FVIII is the coagulation factor deficient in the X-chromosome-linked bleeding disorder haemophilia A. Several other mutant FVIII proteins have also been created (see W33222-29). The FVIII mutant F3095 (W33225) is capable of recombinant secretion at higher levels than typically obtained with wild type FVIII and retains pro-coagulant activity. The FVIII mutant R3361 (W33222) and R562K (W33223) are resistant to activated protein C (APC) cleavage. The present FVIII mutant can form a more stable configuration, and have an approximate 5-fold increase in specific activity compared to purified wild type FVIII, while increasing their binding affinity to von Willebrand factor improves their stability. The FVIII proteins can be administered to haemophiliacs, i.e. FVIII replacement therapy, while the nucleic acid molecule can be used for gene therapy.

note: this sequence does not appear in the specification; it was created using sequences from the given references.

Sequence 1383 AA:

Query Match 94.4%; Score 7362.5; DB 1; Length 1383;
Best Local Similarity 96.1%; Pred. No. 0;
Matches 1382; Conservative 0; Mismatches 1; Indels 55; Gaps 1;

```

Oy 20 ATRRYLGAVELSMQSDLGELPVDAFRPPRPKSPFENTSVYKKTLFVEFTDHLFN 79
Db 1 ATRRYLGAVELSMQSDLGELPVDAFRPPRPKSPFENTSVYKKTLFVEFTDHLFN 60
Oy 80 IAKPRPMGLGPTIOAEVYDTVVITLKNASHPVSLHAGVSYMKASEGAEYDQTSQ 139
Db 61 IAKPRPMGLGPTIOAEVYDTVVITLKNASHPVSLHAGVSYMKASEGAEYDQTSQ 120
Oy 140 REKEDKVPFGSGSHYVQVLEKNGPMASDPLCTLYSHVDLYKDLNSGLIGALLVCR 199
Db 121 REKEDKVPFGSGSHYVQVLEKNGPMASDPLCTLYSHVDLYKDLNSGLIGALLVCR 180
Oy 200 EGSLLAKEQTOLHFKFILFAVFEDEKSMHSETKNSLMODRDAASRAAMPKMTVGYNR 259
Db 181 EGSLLAKEQTOLHFKFILFAVFEDEKSMHSETKNSLMODRDAASRAAMPKMTVGYNR 240
Oy 260 SLPLGLGHRKSVYVHVGMTTPEVHSIFLEGHTFLVNRHROASLESPITFLTAQTL 319
Db 241 SLPLGLGHRKSVYVHVGMTTPEVHSIFLEGHTFLVNRHROASLESPITFLTAQTL 300
Oy 320 MDLGGLFLFCHTSHQHDMGEAYVVDSCPEBPQLRMKNNEAEYDDDLTDEMDVYRF 379
Db 301 MDLGGLFLFCHTSHQHDMGEAYVVDSCPEBPQLRMKNNEAEYDDDLTDEMDVYRF 360
Oy 380 DDDNSPSTQIRSVAKKPKTWVHTIAAEEEMWDYAPLVLAADDSDSYKQYLNNGPORG 439
Db 361 DDDNSPSTQIRSVAKKPKTWVHTIAAEEEMWDYAPLVLAADDSDSYKQYLNNGPORG 420
Oy 440 RRYKVRPMAYTDEFTKREAIQHSGLIGPLLYGEGVDTLLIFKNQASRPYNTYPHGI 499
Db 421 RRYKVRPMAYTDEFTKREAIQHSGLIGPLLYGEGVDTLLIFKNQASRPYNTYPHGI 480
Oy 500 TDVRLPLYSRRLPKYKHLKDFPLPGELFKYKWTYVEDGPKSPDCLTRYSSFVNE 559
Db 481 TDVRLPLYSRRLPKYKHLKDFPLPGELFKYKWTYVEDGPKSPDCLTRYSSFVNE 540
Oy 560 RDLASGLIGPLLYCKESVDQNGQIMSDKRNVLFSVDEKRSMTLTENIORFLPNPAG 619
Db 541 RDLASGLIGPLLYCKESVDQNGQIMSDKRNVLFSVDEKRSMTLTENIORFLPNPAG 600
Oy 620 VOLEDEPEQASIMHSINGVEDSLQSVCLHEVAYWYLLSIGAOTDELFVSFGYTEKH 679
Db 601 VOLEDEPEQASIMHSINGVEDSLQSVCLHEVAYWYLLSIGAOTDELFVSFGYTEKH 660
Oy 680 KAVYEDTLTLFFPSGSETVMSMENPGMLTLCGHSNDFRNKGTALLKYSKCDKNTGDYE 739
Db 661 KAVYEDTLTLFFPSGSETVMSMENPGMLTLCGHSNDFRNKGTALLKYSKCDKNTGDYE 720
Oy 740 DSYEDISATLAKNNNAIEFRSFSQNPVLAKRQRETRTTLSDQDEIDYDITISVEMKK 799

```

```

Db 721 DSYEDISATLAKNNNAIEFRSFSQNPVLAKRQRETRTTLSDQDEIDYDITISVEMKK 739
Oy 800 EDFDIYDEENOSPRSFQKTRHYFIAAVERLMDYGMSSSPHYLRNRAQSGSVQFKRV 859
Db 729 -----PASFOKTRHYFIAAVERLMDYGMSSSPHYLRNRAQSGSVQFKRV 785
Oy 860 FOEFIDGFTQPLRGELNEHLGLGPTIRAEVEDNINWTRNQASRPYSYSLISTEE 919
Db 786 FOEFIDGFTQPLRGELNEHLGLGPTIRAEVEDNINWTRNQASRPYSYSLISTEE 845
Oy 920 DOROGAEPRKFFVNPENKTFFMKVOHMAPTKDEFDKAAVSDVDLEVDHSGLLGP 979
Db 846 DOROGAEPRKFFVNPENKTFFMKVOHMAPTKDEFDKAAVSDVDLEVDHSGLLGP 905
Oy 980 LLVCHTNLNPAGHROVYVQEFALFTTFDETKSVYFENNERNCRAPCNTOMEDPTFE 1039
Db 906 LLVCHTNLNPAGHROVYVQEFALFTTFDETKSVYFENNERNCRAPCNTOMEDPTFE 965
Oy 1040 NYRFAINGYTMDLPLGLVMAQDQIRRYLLSKGSNENHSHIRSGHYFTRKKEEYMA 1099
Db 966 NYRFAINGYTMDLPLGLVMAQDQIRRYLLSKGSNENHSHIRSGHYFTRKKEEYMA 1025
Oy 1100 LYNLPQGVETVEMLPKAGIMRVCECLGEHLHAGMSTLFYVSNKCTPIGMASGHIRD 1159
Db 1026 LYNLPQGVETVEMLPKAGIMRVCECLGEHLHAGMSTLFYVSNKCTPIGMASGHIRD 1085
Oy 1160 FOITASGOYGMAPRLARLHSGSINAMSTKEPSWIKVDLAPWINGITOGAROKFS 1219
Db 1086 FOITASGOYGMAPRLARLHSGSINAMSTKEPSWIKVDLAPWINGITOGAROKFS 1145
Oy 1220 SLTYSQFLIMTSLDGKKQOTYRGNSGTLMVFEQVNDSSGJKHNINFPITARIHLPT 1279
Db 1146 SLTYSQFLIMTSLDGKKQOTYRGNSGTLMVFEQVNDSSGJKHNINFPITARIHLPT 1205
Oy 1280 HYSIRSTRLMELMGCDLNSCMPGMEKASIDQITASSFTNMFATWSPSKARLHOG 1339
Db 1206 HYSIRSTRLMELMGCDLNSCMPGMEKASIDQITASSFTNMFATWSPSKARLHOG 1265
Oy 1340 RSNAMRPVNNPKEMLQVDFQRTMKVGTGTQGVKSLTYSYVKEFLISSQDGHQWTLF 1399
Db 1266 RSNAMRPVNNPKEMLQVDFQRTMKVGTGTQGVKSLTYSYVKEFLISSQDGHQWTLF 1325
Oy 1400 FONGKVKYFQGNQDSFTYVNSLDPPLTRYLRLHPQSWHQIALRMVLCCEADLY 1457
Db 1326 FONGKVKYFQGNQDSFTYVNSLDPPLTRYLRLHPQSWHQIALRMVLCCEADLY 1383

RESULT 11
W33229
ID W33229 standard; protein; 1383 AA.
AC W33229;
DE 30-APR-1998 (first entry)
DT Procoagulant-active human factor VIII:C (FVIII) mutant protein.
KW pro-coagulant active factor VIII; FVIII; haemophilia A;
KW recombinant secretion; pro-coagulant activity; resistance;
KW activated protein C cleavage; APC; B domain; A2 domain; A3 domain;
KW von Willebrand factor binding site; binding affinity;
KW FVIII replacement therapy.
OS Synthetic.
OS Homo sapiens.
FH Key
FH 1. 346 Location/Qualifiers
FT Region
FT /note= "factor VIIIA heavy chain"
FT 741. 1383
FT /note= "factor VIIIA light chain"
FT 1. 329
FT Domain
FT /note= "A1 domain"
FT 1. 179
FT Domain
FT /note= "plastocyanin-like domain 1"
FT 187. 329
FT /note= "plastocyanin-like domain 2"
FT 380. 711
FT Domain

```


FT Misc_feature /note="A2 domain"
 FT 711..746
 FT /note="a spacer of the sequence
 FT SEFQNSRSPSTROKOFNATITIPENDIKTPWF
 FT AHRTPMKIONVSSDLMLL is inserted
 FT between domains A2 and A3"
 FT
 FT Domain 380..554
 FT /note="Plastocyanin-like domain 3"
 FT 564..711
 FT /note="Plastocyanin-like domain 4"
 FT 746..1073
 FT /note="A3 domain"
 FT 1073..1221
 FT /note="C1 domain"
 FT 1226..1378
 FT /note="C2 domain"
 FT 372..373
 FT /note="by thrombin"
 FT Disulfide_bond 153..179
 FT /note="probable"
 FT Disulfide_bond 528..554
 FT /note="probable"
 FT Misc_difference 309
 FT /label="F309S
 FT /note="wild type phe replaced with Ser"
 FT Misc_difference 740
 FT /label="R740A
 FT /note="wild type Arg replaced with Ala"
 FT
 FT MO9740145-A1.
 PD 30-OCT-1997.
 PR 24-APR-1997: U06563.
 PR 15-MAY-1996: US-017785.
 PR 24-APR-1996: US-016117.
 PA (UNM1) UNIV MICHIGAN.
 PI Amano K, Kaufman RJ, Pipe SW;
 DR WPI: 97-535830/49.
 PT Modified human pro-coagulant active factor VIII - can be
 PT administered to haemophiliacs, i.e. factor VIII replacement therapy
 PS Claim 19: Page -: 57pp: English.
 CC The present sequence represents a novel pro-coagulant active factor
 CC VIII (FVIII) mutant protein, comprising a deletion of the B domain and
 CC von Willebrand factor binding site, mutations F309S, R740A and addition
 CC of an amino acid sequence spacer between the A2 and A3 domains. Factor
 CC VIII, along with calcium and phospholipid, acts as a cofactor for factor
 CC IXA, when it converts factor X to the activated form (factor XA). FVIII
 CC is the coagulation factor deficient in the X-chromosome-linked bleeding
 CC disorder haemophilia A. Several other mutant FVIII proteins have also
 CC been created (see W33222-29). The FVIII mutant F309S (W33225) is capable
 CC of recombinant secretion at higher levels than typically obtained with
 CC R3361 (W33222) and R562K (W33223) are resistant to activated protein
 CC C (APC) cleavage. The present FVIII mutant can form a more stable
 CC configuration, and have an approximate 5-fold increase in specific
 CC activity compared to purified wild type FVIII, while increasing their
 CC binding affinity to von Willebrand factor improves their stability.
 CC The FVIII proteins can be administered to haemophiliacs, i.e. FVIII
 CC replacement therapy, while the nucleic acid molecule can be used for
 CC gene therapy.
 CC note: this sequence does not appear in the specification; it was created
 CC using sequences from the given references.
 CC Sequence 1383 AA;
 SQ
 Query Match 94.3%; Score 7354.5; DB 1; Length 1383;
 Best Local Similarity 96.0%; Pred. No. 0;
 Matches 181; Conservative 0; Mismatches 2; Indels 55; Gaps 1;
 QY 20 ATRRRYLGAVELSDMYQSDLGELPYDARFPPRPVPSSEFNTSVYKKTLEVEFTDHLFN 79
 DB 1 ATRRRYLGAVELSDMYQSDLGELPYDARFPPRPVPSSEFNTSVYKKTLEVEFTDHLFN 60
 QY 80 IAKRPPMGLGPTIOAEVYDVTVITLKNMASHPVSLAAGVSYWKASGAEYDQTSQ 139
 DB IAKRPPMGLGPTIOAEVYDVTVITLKNMASHPVSLAAGVSYWKASGAEYDQTSQ 139

Db 61 IAKRPPMGLGPTIOAEVYDVTVITLKNMASHPVSLAAGVSYWKASGAEYDQTSQ 120
 QY 140 REKEDDKYFPGSGHTYVQVYLKENGPAASDPLCTLYSLHYDLYKDLNSGLIGALLYCR 199
 Db 121 REKEDDKYFPGSGHTYVQVYLKENGPAASDPLCTLYSLHYDLYKDLNSGLIGALLYCR 180
 QY 200 EGSIAKEKQTILHKKFILLFAVDEGKSMHSETKNSLMQDRDASARAMPKMTVNGYVNR 259
 Db 181 EGSIAKEKQTILHKKFILLFAVDEGKSMHSETKNSLMQDRDASARAMPKMTVNGYVNR 240
 QY 260 SLPLGICGHRKSYVHWYVGMGTPEVHSIFLEGHTFLVRNRQASLEISPTFTLTAOTLL 319
 Db 241 SLPLGICGHRKSYVHWYVGMGTPEVHSIFLEGHTFLVRNRQASLEISPTFTLTAOTLL 300
 QY 320 MDLGGFLFCHSHSHQHGMEAYVYKVDSCPEPOLRMKNNEADYDDDLTDEMDYVRF 379
 Db 301 MDLGGFLFCHSHSHQHGMEAYVYKVDSCPEPOLRMKNNEADYDDDLTDEMDYVRF 360
 QY 380 DDDNSPFIQIRSVAKKHPKTVWHYIAAEEDWDYAPLVLPDDRSTKSQYLANGPORIG 439
 Db 361 DDDNSPFIQIRSVAKKHPKTVWHYIAAEEDWDYAPLVLPDDRSTKSQYLANGPORIG 420
 QY 440 RYKKVYRPMAYTDEFKTRREALQHSGLIGPLTGEVDTLLIFKNOASRPYNIYPHGI 499
 Db 421 RYKKVYRPMAYTDEFKTRREALQHSGLIGPLTGEVDTLLIFKNOASRPYNIYPHGI 480
 QY 500 TDVRPLYSRLPKGVKHLKDFILPEGEIFKKYKWTVEDEGPTKSDPRLTRYSSFVME 559
 Db 481 TDVRPLYSRLPKGVKHLKDFILPEGEIFKKYKWTVEDEGPTKSDPRLTRYSSFVME 540
 QY 560 RDLASGLIGPLLCYKESVDORGNQIMSDKRNVLIFSVDENKSYLLENQRFELPNAG 619
 Db 541 RDLASGLIGPLLCYKESVDORGNQIMSDKRNVLIFSVDENKSYLLENQRFELPNAG 600
 QY 620 VOLEDEPQASNMHSINQVYFDSQLSVCLHEVAYWYILSGAOTDLSVFFSGYTEKH 679
 Db 601 VOLEDEPQASNMHSINQVYFDSQLSVCLHEVAYWYILSGAOTDLSVFFSGYTEKH 660
 QY 680 KMYEDTLTLFPFSGETVMSMENPGLWILGCHNSDFNRNGTALLKYSSCDKNTGYIE 739
 Db 661 KMYEDTLTLFPFSGETVMSMENPGLWILGCHNSDFNRNGTALLKYSSCDKNTGYIE 720
 QY 740 DSYEDISATILSNNAIEPRSSQNPVLAKRRORETRITLQSDGEIDYDITISVEMKK 799
 Db 721 DSYEDISATILSNNAIEPRSSQNPVLAKRRORETRITLQSDGEIDYDITISVEMKK 739
 QY 800 EDEDIYDEENQSPRSFQKTRHYFLAVERLMDYGSSPHVLRNRAQSGSVPOFKKYV 859
 Db 739 EDEDIYDEENQSPRSFQKTRHYFLAVERLMDYGSSPHVLRNRAQSGSVPOFKKYV 785
 QY 860 FOEFTDGSFTQPLRYKELNENHGLGLPTRYRAVEYEDNIMTFNRQASRPYSFSLISYBE 919
 Db 786 FOEFTDGSFTQPLRYKELNENHGLGLPTRYRAVEYEDNIMTFNRQASRPYSFSLISYBE 845
 QY 920 DOROGAEPKRNFKPNETKITFEKVOHNAAPTDEEDCKAMAFYSVDYDEKDVHSLIGP 979
 Db 846 DOROGAEPKRNFKPNETKITFEKVOHNAAPTDEEDCKAMAFYSVDYDEKDVHSLIGP 905
 QY 980 LLYOCHTILNPAHGRQYVQEFALFTIDETKSWYFTENMERNCAPCNIQMEDPTFE 1039
 Db 906 LLYOCHTILNPAHGRQYVQEFALFTIDETKSWYFTENMERNCAPCNIQMEDPTFE 965
 QY 1040 NYRFAHNGYIMDTLPELVNAODORIRWYLLSMGSNNHSHIFSGHVFVVRKKEEYKMA 1099
 Db 966 NYRFAHNGYIMDTLPELVNAODORIRWYLLSMGSNNHSHIFSGHVFVVRKKEEYKMA 1025
 QY 1100 LYNLYPGVFETVEKLPKAGIWRRECLIGELHLAGNSTLFLVYSNKCQPTLGNASGHIRD 1159
 Db 1026 LYNLYPGVFETVEKLPKAGIWRRECLIGELHLAGNSTLFLVYSNKCQPTLGNASGHIRD 1085
 QY 1160 FOITASQOYGOMAPKLARLHYSGSINAMSTKEPSSATKVDLAPMITHGIKTQGAROKFS 1219
 Db 1086 FOITASQOYGOMAPKLARLHYSGSINAMSTKEPSSATKVDLAPMITHGIKTQGAROKFS 1145


```

Db 421 RYKXKVMATYDETFKREAIQHSGLIGLLYGEVDTLLIIFKMQASRPYNIYPHGI 480
QY 500 TDVRLYSLRRLPKGYKHLKDFILPGELFKYKWTVEDGPKSPDCLTRYSSFFVME 559
Db 481 TDVRLYSLRRLPKGYKHLKDFILPGELFKYKWTVEDGPKSPDCLTRYSSFFVME 540
QY 560 RDLASGLIGLLICKEVDQGNQIMSDRNVILFSEVDENRSWYLTENTIORFLPNPAG 619
Db 541 RDLASGLIGLLICKEVDQGNQIMSDRNVILFSEVDENRSWYLTENTIORFLPNPAG 600
QY 620 VQLEPPEEQASIMHSINGYVDSIQLSVCLHEVAYWYLSIGAOTDPLSVFSGYTEKH 679
Db 601 VQLEPPEEQASIMHSINGYVDSIQLSVCLHEVAYWYLSIGAOTDPLSVFSGYTEKH 660
QY 680 KMYVEDTLTFEFGSEFYMSENMENGLMWLGCHNSDFNRGRTALIKYSSCDKNGDYXE 739
Db 661 KMYVEDTLTFEFGSEFYMSENMENGLMWLGCHNSDFNRGRTALIKYSSCDKNGDYXE 720
QY 740 DSYEDISAYLLSKNNAIEPRFSQNPVYLRHQRHREITRTTLOSDQDEIDYDDTISVEMKK 799
Db 721 DSYEDISAYLLSKNNAIE----- 739
QY 800 EDFDIYDDENQSPRSPQKTRHYFLAVERLMDYGMSSPHVLNRRAQSGSVPOFKKYV 859
Db 739 -----PASFOKTRHYFLAVERLMDYGMSSPHVLNRRAQSGSVPOFKKYV 785
QY 860 FOEFTDGSFTQPLRYGELNEHLGLGPYIRAEVEDNIWYTFENQASRPYSFSSISYEE 919
Db 786 FOEFTDGSFTQPLRYGELNEHLGLGPYIRAEVEDNIWYTFENQASRPYSFSSISYEE 845
QY 920 DQOGAEPKRNKFNENETKYTFWKVQHMAPTKDFEDCKAMAYFSDVDLEKDVHSGLIGP 979
Db 846 DQOGAEPKRNKFNENETKYTFWKVQHMAPTKDFEDCKAMAYFSDVDLEKDVHSGLIGP 905
QY 980 LLYCHTNLNPAGHGVMPVOEALFEFTJFDETKSWYTFENMRNRAPCNIDMEPTKE 1039
Db 906 LLYCHTNLNPAGHGVMPVOEALFEFTJFDETKSWYTFENMRNRAPCNIDMEPTKE 965
QY 1040 NYRFAINGYIMDTLPGLVMAQDORIRWYLSMGSNENHSHFSGHYTFVKKKEEYKA 1099
Db 966 NYRFAINGYIMDTLPGLVMAQDORIRWYLSMGSNENHSHFSGHYTFVKKKEEYKA 1025
QY 1100 LYNLYPGYFVEYEMLPKSAKIRVECLIGEHLHAGMSTFLVYSNCKQTPGLMASGHIRD 1159
Db 1026 LYNLYPGYFVEYEMLPKSAKIRVECLIGEHLHAGMSTFLVYSNCKQTPGLMASGHIRD 1085
QY 1160 FOITASGOYGOWAPKLARLHSGSINAMSTKEPFSMIKVDLAPMTHIKIKQAGROKFS 1219
Db 1086 FOITASGOYGOWAPKLARLHSGSINAMSTKEPFSMIKVDLAPMTHIKIKQAGROKFS 1145
QY 1220 SLYISQFTIMYSLDGKMQOTYRGNSTGTLMVFFGNVDSGSIKHNIFNPPIIARYIRLHPT 1279
Db 1146 SLYISQFTIMYSLDGKMQOTYRGNSTGTLMVFFGNVDSGSIKHNIFNPPIIARYIRLHPT 1205
QY 1280 HYSINSTLMEIMGCDLNSCSPMLGMEKASIDAOITASSYTNMFATWSPKARLHDOG 1339
Db 1206 HYSINSTLMEIMGCDLNSCSPMLGMEKASIDAOITASSYTNMFATWSPKARLHDOG 1265
QY 1340 RSNMARPPOVNNPKEMLOVDFQTKMKVGTGTQGVKSILTSMTVKEFLISSSDGHWITLF 1399
Db 1266 RSNMARPPOVNNPKEMLOVDFQTKMKVGTGTQGVKSILTSMTVKEFLISSSDGHWITLF 1325
QY 1400 FONGKRVYKFGQNSDFTPVVNSLDPPLRLRYLRHPOSVWNOIALRMEVLGECAEADLY 1457
Db 1326 FONGKRVYKFGQNSDFTPVVNSLDPPLRLRYLRHPOSVWNOIALRMEVLGECAEADLY 1383

```

RESULT 13
 W11422
 ID W11422 standard; Protein; 2342 AA.
 AC W11422:
 DT 20-NOV-1997 (first entry)
 NE Active Factor VIII:C analogue, delta 1311-1320, + Pro insertion.

```

KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
KW proteolytic cleavage.
OS Homo sapiens.
OS Synthetic.
FH Key
FH peptide
FH 1..19
FH /note= "signal peptide"
FH protein
FH 20..2342
FH /note= "mature Factor VIII:C"
FH region
FH 20..1658
FH /note= "heavy chain fragment"
FH misc-difference
FH 1329..1330
FH /note= "site of 10 residue deletion"
FH misc-difference
FH 1330
FH /note= "inserted residue"
FH region
FH 1659..2341
FH /note= "light chain fragment"
FH domain
FH 760..1658
FH /note= "B domain"
PN MO9703195-A1.
PD 30-JAN-1997.
PF 09-JUL-1996; U11444.
PR 11-JUL-1995; US-001025.
PA (CHIR ) CHIRON CORP.
PI Cohen FE, Hung DT, Innis M;
DR WPI: 97-119050/11.
PT Factor VIII:C analog modified adjacent to a non-activating Arg
PT residue - used in the treatment of haemophilias, by improvement of
PT haemostasis
PS Claim 27; Page -: 90pp; English.
CC W11330-W11472 represent active Factor VIII:C analogues of the invention.
CC These sequences were created by mutating the wild type Factor VIII:C
CC coding sequence (see J51357) using mutagenic primers. The analogues
CC comprise a native Factor VIII:C polypeptide modified at a site adjacent
CC to a non-activating Arg residue so that a Arg-Pro or Pro-Arg dipeptide is
CC created. Factor VIII:C is a large glycoprotein that participates in the
CC blood coagulation cascade that ultimately converts soluble fibrinogen to
CC insoluble fibrin clot, effecting haemostasis. A deficiency in Factor
CC VIII:C is responsible for haemophilia A, which is an X-chromosome-linked
CC inherited bleeding diathesis. Factor VIII:C is activated by plasma
CC proteases, such as thrombin. During activation the mature polypeptide is
CC cleaved to generate heavy and light chain fragments that are further
CC cleaved. Complexes of two or more of the analogues, nucleic acids and
CC vectors encoding them may be used alone or in conjunction with each
CC other, for the prevention or treatment of active Factor VIII:C deficiency
CC in a mammal. The analogues may be used as immunogens to raise antibodies,
CC and in the treatment of haemophilias, by improvement of haemostasis. The
CC analogues are resistant to proteolytic cleavage and display increased
CC plasma half-life. They may be administered at lower dosages and by
CC different modes of administration.
SQ Sequence 2342 AA:

```

Query Match 94.2%; Score 7344.5; DB 1; Length 2342;
 Best Local Similarity 62.2%; Pred. No. 0;
 Matches 1457; Conservative 0; Mismatches 0; Indels 885; Gaps 1;

```

QY 1 MQIELSTFFCLLRFCSATRRYILGAVELSDMYQSDGLGVVDARFPRVRSKSPFN 60
Db 1 MQIELSTFFCLLRFCSATRRYILGAVELSDMYQSDGLGVVDARFPRVRSKSPFN 60
QY 61 TSYYKKTFLFEFDHLENIAKPRPPMGLGPTIOAEVYTVYITLKNMASHPSLHAY 120
Db 61 TSYYKKTFLFEFDHLENIAKPRPPMGLGPTIOAEVYTVYITLKNMASHPSLHAY 120
QY 121 GVSYWKASGAEDDQTSQREKEDKVPFGSGSHYVQVLKENGPMASDPLCLTYSTLSH 180
Db 121 GVSYWKASGAEDDQTSQREKEDKVPFGSGSHYVQVLKENGPMASDPLCLTYSTLSH 180
QY 181 VDLVKDLNSGLIGALLVCRESLAKETQTLAKFTLLFAVDEGKSWHSEFKNSLMDQRD 240
Db 181 VDLVKDLNSGLIGALLVCRESLAKETQTLAKFTLLFAVDEGKSWHSEFKNSLMDQRD 240

```

Db	181	VDLYKDLNSGLIGALLVCRCGSLAKERTQTLHKFIILFVAFVDEBGSWMSETRKNSLMDROD	240
Qy	241	AASRAAMPKMHYNGVYNSRLPGLICGCHRSYVMHYIGMGTPEVHSLFLEBGTFLVRNH	300
Db	241	AASRAAMPKMHYNGVYNSRLPGLICGCHRSYVMHYIGMGTPEVHSLFLEBGTFLVRNH	300
Qy	301	ROASLESPITFLAOTLLMDLQGFLLFCHSHSHQHDGEAVKVDSCPEEPOLRMKNE	360
Db	301	ROASLESPITFLAOTLLMDLQGFLLFCHSHSHQHDGEAVKVDSCPEEPOLRMKNE	360
Qy	361	EAEYDDDLTDSEMDVYRFDDNSPSFIOIRSAKHKHPWTWHYTAABEEDMDVAPLVLA	420
Db	361	EAEYDDDLTDSEMDVYRFDDNSPSFIOIRSAKHKHPWTWHYTAABEEDMDVAPLVLA	420
Qy	421	PDDRSYSOYLNNGPORIGRKKYKRVFMAVYDETFEKTREAIQHESGIIGPLLYGCVGDTL	480
Db	421	PDDRSYSOYLNNGPORIGRKKYKRVFMAVYDETFEKTREAIQHESGIIGPLLYGCVGDTL	480
Qy	481	LIIFKNOASRPYNIYPHGITDVAPRLYSRRLPGVKHKLDFILPGEIIFKRYKWYTVEDGP	540
Db	481	LIIFKNOASRPYNIYPHGITDVAPRLYSRRLPGVKHKLDFILPGEIIFKRYKWYTVEDGP	540
Qy	541	TKSDPCLRTYSSFVMMERDLASGLIGPLLCYKESVQNRQNOJMSKRVNILLSVDE	600
Db	541	TKSDPCLRTYSSFVMMERDLASGLIGPLLCYKESVQNRQNOJMSKRVNILLSVDE	600
Qy	601	NRSWYLTENIOERLPNAPGQLEDPEFOASNMHISINGVDFSLDASVCLHEVAYWYLLS	660
Db	601	NRSWYLTENIOERLPNAPGQLEDPEFOASNMHISINGVDFSLDASVCLHEVAYWYLLS	660
Qy	661	IGAOTDFLSVFEQGYTFKHKMYEDTLTLPPESGETVEMENPGLMILGCHNSDFRNG	720
Db	661	IGAOTDFLSVFEQGYTFKHKMYEDTLTLPPESGETVEMENPGLMILGCHNSDFRNG	720
Qy	721	MTALLKXSSODKMGDYEDSYEDISAYLLSKKNALPEPFSF-----	762
Db	721	MTALLKXSSODKMGDYEDSYEDISAYLLSKKNALPEPFSFSONSRHPSTRKOFNAATTI	780
Qy	762	-----	762
Db	781	PENDIEKTDWFAHRTMPRIQWVSSDLLMLLRQSPTRHGSLSDLDKATYETPSDDPS	840
Qy	762	-----	762
Db	841	PGAIDSNNSLSEMTHFRPQLHSGDMVETPESGLQRLNEKLGTTAAATELKKDEKYST	900
Qy	762	-----	762
Db	901	SNNLISTIPSDNLAAGTDNNTSLGPPSPMYHYDSOLDITTLFGKKSSPLTESGGLSSEE	960
Db	762	-----	762
Db	961	NNDSKLLSGLMSQESBSMCKNWSTESGRLEFKGKRAHPALLTQDNALFKVYSISLKTN	1020
Qy	762	-----	762
Db	1021	KTSNNSATNRKTHIDGSPSLIENSPLYQWNIIESOTEFKKYTPPLIHDRMLDKNATAURL	1080
Qy	762	-----	762
Db	1081	NHMSNKTTSKMMEMVOQKKEGPIPPDAQNPDMSPFKMLFLPESARWIOPTHGKNSLSNG	1140
Qy	762	-----	762
Db	1141	QGPSPKQULVSLGPEKSYEGONFLSEKKNKYVVGGETTKDYGLKENVPSSRNLPLTINDN	1200
Qy	762	-----	762
Db	1201	LHENNTHQOEKKIOEIEKKETLIEOENVVLPOIHVYTGAKNFMKMLPLSTHQNVESGD	1260
Qy	762	-----	762
Db	1261	GAYAPVLODFRSLNDSTNRKTKHTAHFSKKKEEENLEGNQTKOIVEKACTTRISENT	1320

QY	762	-----	762
Db	1331	SQONFVQRPPELEETELEKRI IVDJTSIQMSKNMHLAPSTLTQIDYNEKEKAIITOSPL	1380
QY	762	-----	762
Db	1381	SDCLTRSHSIPQANRSLPLAKYSFPPSIRPIYTLFVLEFDNSSHLPAASTRKKDSQVE	1440
QY	762	-----	762
Db	1441	SSHFLQAKKNMLSLALILEMTGDQREVSGLSGTSATNSVTYKKEVNTVLPKPDLPKTS	1500
QY	762	-----	762
Db	1501	KVELLPKVHIYQXDLPEPTETSNQSPGHLDLVEGSLLOGTEGAIKKNANRPGKVPFLRA	1560
QY	762	-----	762
Db	1561	TESSAKTPSKILDLPLANDNHYGTQIPREKMSOESKPEKTAERKKDTLLSLNACESNHA	1620
QY	762	-----SONPVYLKRHOREITRTTQSDOEIEDYDRTISV	795
Db	1621	AINENGONKPELEVTMAKQOETERLCSQNPVYLKRQREITRTTQSDOEIEDYDRTISV	1680
QY	796	EMKKEDPIDYEDENOSPSRFQKTRHYFLAAVERLMDYGNSSPHVLNRAQSGSVQF	855
Db	1681	EMKKEDPIDYEDENOSPSRFQKTRHYFLAAVERLMDYGNSSPHVLNRAQSGSVQF	1740
QY	856	KKVYQOETOSFTQPLXRELENEHGLGPRYRAVEDNIMVTRNQAAPRYSFYSLL	915
Db	1741	KKVYQOETOSFTQPLXRELENEHGLGPRYRAVEDNIMVTRNQAAPRYSFYSLL	1800
QY	916	SYEEDROGAEPRAKNFKPNETKTYFPMKVQNHNAPTKDEFDCKANAAYSDDVLEKDVHSG	975
Db	1801	SYEEDROGAEPRAKNFKPNETKTYFPMKVQNHNAPTKDEFDCKANAAYSDDVLEKDVHSG	1860
QY	976	LIGPLLCHTNTLNPAGROVYQVOEFALFFTJFDETKSMYFTEENNERCRAPCNIOMEDP	1035
Db	1861	LIGPLLCHTNTLNPAGROVYQVOEFALFFTJFDETKSMYFTEENNERCRAPCNIOMEDP	1920
QY	1036	TFKEKYRPHALNGIMNTLFGLYMAODQRIKRYLLMSGNENIHSIHSGHYFYRKKEE	1095
Db	1921	TFKEKYRPHALNGIMNTLFGLYMAODQRIKRYLLMSGNENIHSIHSGHYFYRKKEE	1980
QY	1096	KYKALYNYPGVEFTYMLSKAGIRVBEJLGEHHAHMSLFLYVNSKOTPIGMASS	1155
Db	1981	KYKALYNYPGVEFTYMLSKAGIRVBEJLGEHHAHMSLFLYVNSKOTPIGMASS	2040
QY	1156	HIRDFOLITASGOYQOMAPKLARLHYSGSINAMSTKEPFSMIVYDLAPMIIHGITQGAR	1215
Db	2041	HIRDFOLITASGOYQOMAPKLARLHYSGSINAMSTKEPFSMIVYDLAPMIIHGITQGAR	2100
QY	1216	QKFSLLYISQFIIMYSLDGKKMOTYRGNSTGLTLMVFFGNVDSGKIHNIENPIIARYIR	1275
Db	2101	QKFSLLYISQFIIMYSLDGKKMOTYRGNSTGLTLMVFFGNVDSGKIHNIENPIIARYIR	2160
QY	1276	LAPHTYSTRSLRHELMGCDLNSCNPGLMESKAISDAOITRASSFTTNMFATWSKARL	1335
Db	2161	LAPHTYSTRSLRHELMGCDLNSCNPGLMESKAISDAOITRASSFTTNMFATWSKARL	2220
QY	1336	HLQGSNMAIRQVANNPKEMJQVDFQRTMKYGTVTQGVKSLLTSMYVEKFLTSSSDQHO	1395
Db	2221	HLQGSNMAIRQVANNPKEMJQVDFQRTMKYGTVTQGVKSLLTSMYVEKFLTSSSDQHO	2280
QY	1396	WTLFFQNGKAVFQGNDSFTPVVNSLDPELLTRYLRHPQSVWQIALRMEVLGCEAD	1455
Db	2281	WTLFFQNGKAVFQGNDSFTPVVNSLDPELLTRYLRHPQSVWQIALRMEVLGCEAD	2340
QY	1456	LY 1457	
Db	2341	LY 2342	

Db 1201 LHEHNTHNÖEKKIQÖEIEKKEKTLIQENNVLPQJHTVTGKNFKMKNLFLSTRÖNVESYD 1260
QY 762 ----- 762
Db 1261 GAYAPVLQDERSLNDSNTNRTKHTAHFSKKGEEBLEGLNOTKOIVEKYACTTRISPNT 1320
QY 762 ----- 762
Db 1321 SQONFVTQÖSKRPLEETELEKRIIVDDTSTÖWSKNMHLTPSTLTQIDYNEKEKGAITQS 1380
QY 762 ----- 762
Db 1381 PLSDCLTRSHSIPQANSRPLPAKVSSPSPRIYTLRVLEQDNSSLPAASYKKKDSGV 1440
QY 762 ----- 762
Db 1441 QÖSSHFLQAKKNLNLAILLEMTGQREVGSLGTSATNSVTYKVENYVLPKPLPKT 1500
QY 762 ----- 762
Db 1501 SKGVELLPKVHIYOKDLFPETETSGSPGHLDYEGSLQGTGCAIKWNEANRPGKVPFLR 1560
QY 762 ----- 762
Db 1561 VATESSAKTPSKLUDPLAMDNHYGTQIPKEEMKQÖEKSPEKTAFKKDTLISLNACESNH 1620
QY 762 -----SONPYLAKRHÖREITRTTQSDÖEIDYDITI 793
Db 1621 AIAINEGÖNKPEIEVTWAKÖGRTERLCSÖNPYLPKRRHÖREITRTTQSDÖEIDYDITI 1680
QY 794 SVEKKKEDFIYDEDEÖSPRSFÖKTRHYFIAVERLMYGMSSPHVLRNRQÖSGVP 853
Db 1681 SVEKKKEDFIYDEDEÖSPRSFÖKTRHYFIAVERLMYGMSSPHVLRNRQÖSGVP 1740
QY 854 QFKRVYQÖEFTDGSFTQPLYRGELNEHLGLPYIRAEVEDNINVTFRNOASRPYSYSS 913
Db 1741 QFKRVYQÖEFTDGSFTQPLYRGELNEHLGLPYIRAEVEDNINVTFRNOASRPYSYSS 1800
QY 914 LISIEEÖROGAEPRKRFVFNETKTYFMKVOHMAPTKDEFDCKAMAYFSDVDLEKDVH 973
Db 1801 LISIEEÖROGAEPRKRFVFNETKTYFMKVOHMAPTKDEFDCKAMAYFSDVDLEKDVH 1860
QY 974 SGLIGPLVCHTINTLNAHGRÖVYVÖEALFEFTFDETKSKYFENNERNCRAQCNÖME 1033
Db 1861 SGLIGPLVCHTINTLNAHGRÖVYVÖEALFEFTFDETKSKYFENNERNCRAQCNÖME 1920
QY 1034 DPTFKENYRFHAINGYIMDTLPGLVMAQÖRIRMYILSMGSNENIHSIHFSGHVFTYRKK 1093
Db 1921 DPTFKENYRFHAINGYIMDTLPGLVMAQÖRIRMYILSMGSNENIHSIHFSGHVFTYRKK 1980
QY 1094 EBYKMALYNLYPGVFETVEMLPKRAQIMRWECCLIGEHLHAGMSTLFLVYSNKCOTPLGMA 1153
Db 1981 EBYKMALYNLYPGVFETVEMLPKRAQIMRWECCLIGEHLHAGMSTLFLVYSNKCOTPLGMA 2040
QY 1154 SGRHÖDQIRASQGYGQMAKPLARLHYSGSINASTKEPFSWITVDLAPITIHGITÖG 1213
Db 2041 SGRHÖDQIRASQGYGQMAKPLARLHYSGSINASTKEPFSWITVDLAPITIHGITÖG 2100
QY 1214 AROFSSLYISOFTIMYSLDGKKWQYRGNSTGTLMYFQGVNDSSGJKHNJFNPIIARY 1273
Db 2101 AROFSSLYISOFTIMYSLDGKKWQYRGNSTGTLMYFQGVNDSSGJKHNJFNPIIARY 2160
QY 1274 IRLHPTIYSIRSTIRLMELGCDLNSCMLPGMESKASDAQITASSYFTNNFATWSPSKA 1333
Db 2161 IRLHPTIYSIRSTIRLMELGCDLNSCMLPGMESKASDAQITASSYFTNNFATWSPSKA 2220
QY 1334 RLHÖGSSNARPOVNNPKELQYDFÖKTKMYVTYTTQÖKASLLTSMTYKFEELISSQÖG 1393
Db 2221 RLHÖGSSNARPOVNNPKELQYDFÖKTKMYVTYTTQÖKASLLTSMTYKFEELISSQÖG 2280
QY 1394 HQWTLFQÖNGKVKYFQÖNÖDSFTPVVNSLDPPLTLTRYLRIRHPOSVMHÖIALRMKVLGCEA 1453
Db 2281 HQWTLFQÖNGKVKYFQÖNÖDSFTPVVNSLDPPLTLTRYLRIRHPOSVMHÖIALRMKVLGCEA 2340

QY 1454 QDLY 1457
Db 2341 QDLY 2344
RESULT 15
ID W11410 standard; Protein; 2344 AA.
AC W11410;
DT 20-NOV-1997 (first entry)
DE Active Factor VIII:C analogue; delta 768-775, + Pro insertion.
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic therapy;
OS Homo sapiens.
OS Synthetic.
FH key
FT peptide
FT /note= "signal peptide"
FT protein
FT /note= "mature Factor VIII:C"
FT region
FT /note= "heavy chain fragment"
FT misc-difference
FT /note= "site of 8 residue deletion"
FT misc-difference
FT /note= "inserted residue"
FT region
FT /note= "light chain fragment"
FT domain
FT /note= "B domain"
PN MO9703195-A1.
PD 30-JAN-1997.
PF 09-JUL-1996; U11444.
PR 11-JUL-1995; US-001025.
PA (CHIR) CHIRON CORP.
PI Cohen FE, Hung DT, Innis M;
DR WPI; 97-119050/11.
PT Factor VIII:C analog modified adjacent to a non-activating Arg
PT residue - used in the treatment of haemophilias, by improvement of
PT haemostasis
PS Claim 25; Page -: 90pp; English.
CC W11330-W11472 represent active Factor VIII:C analogues of the invention.
CC These sequences were created by mutating the wild type Factor VIII:C
CC coding sequence (see F51357) using mutagenic primers. The analogues
CC comprise a native Factor VIII:C polypeptide modified at a site adjacent
CC to a non-activating Arg residue so that a Arg-Pro or Pro-Arg dipeptide is
CC created. Factor VIII:C is a large glycoprotein that participates in the
CC blood coagulation cascade that ultimately converts soluble fibrinogen to
CC insoluble fibrin clot, effecting haemostasis. A deficiency in Factor
CC VIII:C is responsible for haemophilia A, which is an X-chromosome-linked
CC inherited bleeding diathesis. Factor VIII:C is activated by plasma
CC proteases, such as thrombin. During activation the mature polypeptide is
CC cleaved to generate heavy and light chain fragments that are further
CC cleaved. Complexes of two or more of the analogues, nucleic acids and
CC vectors encoding them may be used alone or in conjunction with each
CC other, for the prevention or treatment of active Factor VIII:C deficiency
CC in a mammal. The analogues may be used as immunogens to raise antibodies,
CC and in the treatment of haemophilias, by improvement of haemostasis. The
CC analogues are resistant to proteolytic cleavage and display increased
CC plasma half-life. They may be administered at lower dosages and by
CC different modes of administration.
SQ Sequence 2344 AA;
Query Match 94.2%; Score 7343.5; DB 1; Length 2344;
Best Local Similarity 62.2%; Pred. No. 0;
Matches 1457; Conservative 0; Mismatches 0; Indels 887; Gaps 1;
QY 1 MOEFLSCFLCLIRFSSATRRYIGAVELSDMYDQSDGELPVDARFPKSPFN 60
|||||

Db 1 MOELSTGFLCLLRFCSATRRYLGAVELSMOYMSDGLGELPVDARPPRVPKSPFN 60
 Qy 61 TSVVYKKTLEVEETDHLNIAKRPMMGLGPTIOAEVYDTVTTLKNMASHPVSLAV 120
 Db 61 TSVVYKKTLEVEETDHLNIAKRPMMGLGPTIOAEVYDTVTTLKNMASHPVSLAV 120
 Qy 121 GVSYKASGAGYDQOTSOREKEDKVPFGSGHYWOLKENGMSADPLCLTSTYSH 180
 Db 121 GVSYKASGAGYDQOTSOREKEDKVPFGSGHYWOLKENGMSADPLCLTSTYSH 180
 Qy 181 VDLVNDLNGLIGALLVREGSLAKEKTOTLHKLFLFAVEDEGKSWHSETKNSIMODRD 240
 Db 181 VDLVNDLNGLIGALLVREGSLAKEKTOTLHKLFLFAVEDEGKSWHSETKNSIMODRD 240
 Qy 241 AASARAMPMTHTVNGVYVNSLPGLIGCHRRKSYVHWYIGTTPVHSTFLBEHTLYVNH 300
 Db 241 AASARAMPMTHTVNGVYVNSLPGLIGCHRRKSYVHWYIGTTPVHSTFLBEHTLYVNH 300
 Qy 301 RQASLEISPTFLTAQTLMLDGOFLLFCHISSHQHDGMEAYKVYDSCPEEPQLMKKNNE 360
 Db 301 RQASLEISPTFLTAQTLMLDGOFLLFCHISSHQHDGMEAYKVYDSCPEEPQLMKKNNE 360
 Qy 361 EADYDDDLTJDEMDVYRPDDDNPSFQIRSVAKKHPTWHTYIAEEDMDYAPLYLA 420
 Db 361 EADYDDDLTJDEMDVYRPDDDNPSFQIRSVAKKHPTWHTYIAEEDMDYAPLYLA 420
 Qy 421 PDDRSKSYLNNGPORIRKRYKRYREMAVTDTEKTRALIOHESGILLPLLYGEGDTL 480
 Db 421 PDDRSKSYLNNGPORIRKRYKRYREMAVTDTEKTRALIOHESGILLPLLYGEGDTL 480
 Qy 481 LLIFFKQARPNYIYHGTIDVAPLYSRRLPKGVKHLKDFPLPGEIFYKMYTVEDEGP 540
 Db 481 LLIFFKQARPNYIYHGTIDVAPLYSRRLPKGVKHLKDFPLPGEIFYKMYTVEDEGP 540
 Qy 541 TKSDDPCLTRYSYFVNMERDLASGLIGPLLYCKESVQOQNGOIMSDKRNVTLYFVPE 600
 Db 541 TKSDDPCLTRYSYFVNMERDLASGLIGPLLYCKESVQOQNGOIMSDKRNVTLYFVPE 600
 Qy 601 NRSWYLTENTORFLPNPAGVQLEDPEFOASINMHSINGYFDSLOLVCLEHVAAYTILS 660
 Db 601 NRSWYLTENTORFLPNPAGVQLEDPEFOASINMHSINGYFDSLOLVCLEHVAAYTILS 660
 Qy 661 IGAOTDFLSYFSGYTFKHKMYEDTLTLPFSGEYFVMSMNPGLMILGCHNSDFRNR 720
 Db 661 IGAOTDFLSYFSGYTFKHKMYEDTLTLPFSGEYFVMSMNPGLMILGCHNSDFRNR 720
 Qy 721 MTALLKVSODKNTGDYEDSYEDISAYLSKNNALIEPRSF----- 762
 Db 721 MTALLKVSODKNTGDYEDSYEDISAYLSKNNALIEPRSF----- 762
 Qy 762 ----- 762
 Db 762 ----- 762
 Qy 781 PENDIEPRTPMKIONVSSDDLMLLRQSPTHQLSLSDLEAKYETSDPSPGAIDSN 840
 Db 781 PENDIEPRTPMKIONVSSDDLMLLRQSPTHQLSLSDLEAKYETSDPSPGAIDSN 840
 Qy 841 NSISEMTHFRPOLHSGDNYFPESGLOLRNLEKLTAAATELKLDFKVSSTSNLST 900
 Db 841 NSISEMTHFRPOLHSGDNYFPESGLOLRNLEKLTAAATELKLDFKVSSTSNLST 900
 Qy 901 IPSDNLAACTDNTSSLGPPSPMVPYHDSQDLTTLFGKSSPLTESGGPLSSEENNSKLL 960
 Db 901 IPSDNLAACTDNTSSLGPPSPMVPYHDSQDLTTLFGKSSPLTESGGPLSSEENNSKLL 960
 Qy 961 ESGLMNSQESWGKNVSTESGRLFKGKRAHGPALLTKDNALFKVISLTKNTKSNSNA 1020
 Db 961 ESGLMNSQESWGKNVSTESGRLFKGKRAHGPALLTKDNALFKVISLTKNTKSNSNA 1020
 Qy 1021 TNRKTHIDGSLIENSPVQNTLESDETEKVTPLIHDMLMDKNAATLRLNHSNKT 1080
 Db 1021 TNRKTHIDGSLIENSPVQNTLESDETEKVTPLIHDMLMDKNAATLRLNHSNKT 1080
 Qy 1081 TSSKNMEMVOCKEGPTPPAQNPDMSFFKMLFLPESARWLTQTHGKNSLNSGGSPSKO 1140
 Db 1081 TSSKNMEMVOCKEGPTPPAQNPDMSFFKMLFLPESARWLTQTHGKNSLNSGGSPSKO 1140

Qy 762 ----- 762
 Db 1141 LVSIGPEKSEVGNFLSEKKKVVYVGKEFTKDVGLKEMVFPSSRNLETLMDLHENNTH 1200
 Qy 762 ----- 762
 Db 1201 NOEKIOEIEKKEFTLLIOENVVLPQIHTYTGTRNFMKNLEFLSTRONVEGSDGAYAPVL 1260
 Qy 762 ----- 762
 Db 1261 QDFRSLNDSTNRKTKKHAHPSKGEENLEGLNQTYQYVEKACTTRISPNTSQONFV 1320
 Qy 762 ----- 762
 Db 1321 QRSKRALQKPLPLETELEKRIIVDOTSTOWSKMKHLTPSTLTQIDYNEKEGATIOS 1380
 Qy 762 ----- 762
 Db 1381 PLSDCLTRSHSIPQANSPLPIAKVSSPSIRPIYLTIVLFDONSSHLPASTRKDGCV 1440
 Qy 762 ----- 762
 Db 1441 QESHFLQAKKNLSLAILLEMTGDQREVGSIGTSATNSVYKKYENTVLPKPOLPKT 1500
 Qy 762 ----- 762
 Db 1501 SGKVELPKVHIYQDLFPTETNSGPHLDLVEGSLQTEGAIKWNEANRPGKVPFLR 1560
 Qy 762 ----- 762
 Db 1561 VATESAATPSKLLDPLAMDNHYGTQLPKEEMKSOEKSPEKTARKKDDTILSLNACESNH 1620
 Qy 762 ----- 762
 Db 1621 AIAAINEGONKPELEVTAQKQRTERLCSQNPVYLKRQREITRTLOSQDEEIDYDTI 1680
 Qy 762 ----- 762
 Db 1681 SVEMKKEPDFDIYDEENOSPSPSFQKTRHYFIAAVERLMDGMSSPVLRNRRQSGVP 1740
 Qy 762 ----- 762
 Db 1741 QPKKVVFOEFTDGSTQOLYKGLNEHGLGPTIRAVEDINIMVTFENQASRPYSFSS 1800
 Qy 762 ----- 762
 Db 1801 LSYEEDQOGAEPKRNKPVKPRKTYFMKVOHMAPTKDEFDCAKAAYFSDVLEKDVH 1860
 Qy 762 ----- 762
 Db 1861 SGLIGPLVCHNTNLPAGROVTVQEFALFETIDETKSWYFTENMERNCRAPCNIOME 1920
 Qy 762 ----- 762
 Db 1921 DPLFKENYFHAINGYINDTLPLGLVMAQDQRIRWYLLSGMSNENHSHFSGHVFYARK 1980
 Qy 762 ----- 762
 Db 1981 EBYKALVLYGVFEYVMBLPSKAGIMRVECLIEBHLAGSTLELYVSNQCPPLGMA 2040
 Qy 762 ----- 762
 Db 2041 SGHIRFOITTAGOYGOMAPKLARLHSGSINAWSTKEPFSIKYDILAPMIIHIIKTOG 2100
 Qy 762 ----- 762
 Db 2101 ARQKFSILYISOFTIMYSLDGKRWQTYRGNSTGLMVEFGVNDSSGKIHNIENPPIIARY 2160
 Qy 762 ----- 762
 Db 2161 IRLHPTHYISIRSLRMLMGCDLNSGSMPLGMSKRAISDAQTASSYFTNMFATWSPSKA 2220

QY	1334	RLHLQGRSNAMRPOVNNPKEMLOVDFOKTMKVTGVTQGVKSLLTSMYKKEFLISSODG	1393
Db	2221	RLHLQGRSNAMRPOVNNPKEMLOVDFOKTMKVTGVTQGVKSLLTSMYKKEFLISSODG	2280
QY	1394	HOWTLFFQNGKVKVFGQNDSTFPVNSLDPPLLTRYLRIHQSGWVHQIALRMEVLGCEA	1453
Db	2281	HOWTLFFQNGKVKVFGQNDSTFPVNSLDPPLLTRYLRIHQSGWVHQIALRMEVLGCEA	2340
QY	1454	ODLY	1457
Db	2341	QDLY	2344

Search completed: August 19, 1999, 15:59:09
Job time: 2446 sec

